LAPORAN PROJEK 1
IMPLEMENTASI 5 DATASET
MULTIPLE REGRESI LINIER

Dosen Pembimbing:

Farhanna Mar'i, S.Kom., M.Kom.

Disusun oleh:

Muhammad Anwar Ismail (24051204161)

Program Studi Teknik Informatika

Universitas Negeri Surabaya

2025

DAFTAR ISI

DAFTAR ISI **2**

KATA PENGANTAR **5**

BAB I PENDAHULUAN **6**

1.1 Latar Belakang 6

1.2 Rumusan Masalah 6

1.3 Tujuan 7

1.4 Manfaat 7

BAB II KAJIAN PUSTAKA **8**

2.1 Multiple Regresi Linier 8

BAB III PEMBAHASAN **9**

3.1 Dataset 9

3.1.1 Housing Prices Dataset 9

3.1.2 Car Price Dataset 11

3.1.3 Medical Cost Personal Dataset 14

3.1.4 Fish Market 16

3.1.5 3D Printer Dataset 18

3.2 Implementasi Metode Multiple Regresi Linier Terhadap Housing Prices Dataset 21

3.2.1 Tahap Pengerjaan 21

3.2.2 Menyusun Matriks Augmented 21

3.2.3 Transformasi ke Bentuk Segitiga Atas 22

3.2.4 Nilai Error 27

3.2.5 Outlier 28

3.2.6 Source Code 29

3.2.7 Penjelasan Source Code 33

3.2.8 Tampilan Terminal dan Grafik 40
3.3 Implementasi Metode Multiple Regresi Linier Terhadap Car Price Dataset 40
3.3.1 Tahap Pengerjaan.....40
3.3.2 Menyusun Matriks Augmented.....41
3.3.3 Transformasi ke Bentuk Segitiga Atas 42
3.3.4 Nilai Error 46
3.3.5 Outlier 47
3.3.6 Source Code 48
3.3.7 Penjelasan Source Code 52
3.3.8 Tampilan Terminal dan Grafik 58
3.4 Implementasi Metode Multiple Regresi Linier Terhadap Medical Cost Personal Dataset 58
3.4.1 Tahap Pengerjaan.....58
3.4.2 Menyusun Matriks Augmented.....59
3.4.3 Transformasi ke Bentuk Segitiga Atas 60
3.4.4 Nilai Error 64
3.4.5 Outlier 65
3.4.6 Source Code 65
3.4.7 Penjelasan Source Code 69
3.4.8 Tampilan Terminal dan Grafik 75
3.5 Implementasi Metode Multiple Regresi Linier Terhadap Fish Market Dataset ...76
3.5.1 Tahap Pengerjaan.....76
3.5.2 Menyusun Matriks Augmented.....76
3.5.3 Transformasi ke Bentuk Segitiga Atas 77
3.5.4 Nilai Error 81

3.5.5 Outlier 82
3.5.6 Source Code 83
3.5.7 Penjelasan Source Code 87
3.5.8 Tampilan Terminal dan Grafik 93
3.6 Implementasi Metode Multiple Regresi Linier Terhadap 3D Printer Dataset 93
3.6.1 Tahap Pengerjaan 93
3.6.2 Menyusun Matriks Augmented 94
3.6.3 Transformasi ke Bentuk Segitiga Atas 95
3.6.4 Nilai Error 99
3.6.5 Outlier 100
3.6.6 Source Code 101
3.6.7 Penjelasan Source Code 105
3.6.8 Tampilan Terminal dan Grafik 111
3.7 Evaluasi Performa Model 111
BAB IV PENUTUP **114**
4.1 Kesimpulan 114
4.2 Saran 115
DAFTAR PUSTAKA **117**

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan laporan praktikum ini dengan baik. Laporan ini disusun sebagai salah satu tugas mata kuliah Sains Komputasi, dengan judul "Implementasi 5 Dataset Multiple Regresi Linier".

Saya menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, saya mengucapkan terima kasih yang sebesar-besarnya kepada:

- Bapak/Ibu dosen pengampu mata kuliah Sains Komputasi, Ibu Farhanna Mar'i S.Kom., M.Kom., yang telah memberikan bimbingan dan arahan selama praktikum ini berlangsung.
- Rekan-rekan mahasiswa yang turut berdiskusi dan berbagi pengetahuan sehingga laporan ini dapat tersusun dengan baik.

Semoga laporan ini dapat memberikan manfaat, baik sebagai bahan pembelajaran maupun sebagai referensi bagi siapa saja yang ingin memahami lebih dalam tentang implementasi octave. Segala kritik dan saran yang membangun sangat saya harapkan untuk perbaikan laporan ini ke depannya.

Akhir kata, semoga laporan ini dapat diterima dan memberikan gambaran yang jelas mengenai hasil praktikum yang telah saya lakukan.

Surabaya, 8 Oktober 2025

Muhammad Anwar Ismail

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini ditandai dengan melimpahnya data yang dihasilkan dari berbagai sektor kehidupan, seperti ekonomi, kesehatan, pendidikan, dan industri manufaktur. Data tersebut tidak hanya menjadi catatan historis, tetapi juga memiliki potensi besar untuk dianalisis lebih dalam guna memperoleh wawasan berharga. Salah satu pendekatan yang banyak digunakan dalam analisis data adalah regresi linier berganda, yakni sebuah metode statistik untuk memodelkan hubungan linier antara satu variabel dependen dengan beberapa variabel independen.

Sebagai mahasiswa teknik informatika, pemahaman terhadap metode regresi linier berganda tidak hanya penting secara teoritis, tetapi juga aplikatif. Dalam praktikum Sains Komputasi ini, penulis menerapkan regresi linier berganda pada lima dataset berbeda: Housing Prices, Car Price, Medical Cost Personal, Fish Market, dan 3D Printer. Kelima dataset tersebut memiliki karakteristik variabel bebas yang beragam, namun sama-sama bertujuan memprediksi suatu variabel terikat. Melalui projek ini, penulis ingin membuktikan bagaimana regresi linier berganda dapat digunakan untuk memodelkan dan memprediksi data nyata, sekaligus melatih kemampuan analisis dan pengolahan data menggunakan perangkat lunak Octave/MATLAB.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam tugas ini adalah:

1. Bagaimana cara menyusun dan mengolah data dari lima dataset yang berbeda sehingga dapat dianalisis menggunakan regresi linier berganda?
2. Bagaimana membentuk matriks normal equation dan matriks augmented untuk mendapatkan koefisien regresi?
3. Bagaimana proses transformasi matriks augmented menjadi bentuk segitiga atas serta melakukan back-substitution untuk mendapatkan koefisien model?

4. Bagaimana menghitung dan menganalisis nilai error (SSE, MSE, RMSE, MAPE) serta residual untuk masing-masing dataset?
5. Apa saja outlier yang muncul dari setiap model dan bagaimana pengaruhnya terhadap performa prediksi?

1.3 Tujuan

Tujuan tugas ini adalah:

1. Menerapkan algoritma regresi linier berganda pada lima dataset yang berbeda dengan menggunakan bahasa pemrograman Octave/MATLAB.
2. Menyusun matriks normal equation, melakukan eliminasi Gauss, dan back substitution untuk memperoleh koefisien regresi.
3. Menghitung indikator performa model seperti SSE, MSE, RMSE, dan MAPE sehingga dapat menilai seberapa baik model memprediksi data.
4. Mengidentifikasi dan menganalisis outlier yang muncul pada tiap dataset melalui grafik residual.
5. Menarik kesimpulan mengenai keunggulan dan keterbatasan regresi linier berganda dalam memodelkan berbagai jenis data.

1.4 Manfaat

Manfaat yang diharapkan dari tugas ini adalah:

1. Meningkatkan pemahaman mengenai konsep regresi linier berganda, mulai dari teori hingga implementasinya, serta mengasah kemampuan menggunakan perangkat lunak pengolahan data seperti Octave/MATLAB.
2. Memberikan contoh penerapan regresi linier berganda pada beragam dataset yang dapat dijadikan referensi atau studi kasus untuk pembelajaran lanjutan.
3. Menjadi ilustrasi sederhana tentang bagaimana pemilihan variabel bebas memengaruhi performa model prediksi serta pentingnya deteksi outlier.
4. Menyediakan pijakan awal untuk eksplorasi metode statistik lain, seperti regresi non-linier atau machine learning, guna meningkatkan akurasi prediksi pada dataset yang kompleks.

BAB II

KAJIAN PUSTAKA

2.1. Multiple Regresi Linier

Regresi linier berganda adalah suatu metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen. Bentuk umum dari model ini adalah:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \ldots + b_nx_n + e$$

Di mana:

- Y adalah variabel dependen (target),
- $x_1, x_2, \ldots x_n$ adalah variabel independen,
- b_0 adalah intersep,
- $b_1, b_2, \ldots b_n$ adalah koefisien regresi masing-masing variabel independen,
- e adalah galat atau residual dari model.

Tujuan dari regresi ini adalah untuk meminimalkan jumlah kuadrat dari galat prediksi. Salah satu metode populer yang digunakan untuk mendapatkan solusi parameter regresi adalah metode Least Squares. Proses ini melibatkan pembentukan sistem persamaan normal, transformasi ke bentuk matriks segitiga atas melalui eliminasi Gauss, dan penyelesaian menggunakan substitusi mundur.

Implementasi dalam tugas ini menggunakan GNU Octave, sebuah perangkat lunak open-source yang memungkinkan pengguna untuk melakukan perhitungan numerik dan visualisasi data dengan bahasa pemrograman yang mirip dengan MATLAB.

BAB III
PEMBAHASAN

3.1. Dataset

- Sumber dan cakupan data

Lembar kerja olahan: Google Sheets https://docs.google.com/spreadsheets/d/1cqvAicGvf-wpyzHa7CJKMvsIUjkG9aOTDVbMSuxqmOs/edit?usp=sharing

3.1.1. Housing Prices Dataset

Dataset ini terdiri atas 545 observasi dan 13 atribut. Atribut numerik utamanya meliputi harga (price), luas bangunan (area), jumlah kamar tidur (bedrooms), jumlah kamar mandi (bathrooms), jumlah lantai (stories), serta ketersediaan lahan parkir. Selain itu terdapat beberapa atribut kategorikal seperti keberadaan jalan utama (mainroad), kamar tamu (guestroom), ruang bawah tanah (basement), pemanas air (hotwaterheating), pendingin udara (airconditioning), preferensi lokasi (prefarea) dan status furnitur (furnishingstatus). Dataset ini tidak memiliki nilai hilang sehingga siap digunakan tanpa proses imputasi.

Dalam proyek ini, penulis menggunakan 50 data pertama dari dataset tersebut dengan variabel bebas berupa area, bedrooms, bathrooms dan stories, serta variabel terikat price. Relasi yang ingin ditangkap adalah bagaimana ukuran rumah dan jumlah ruang mempengaruhi harga jual.

Sumber Utama: Kaggle - Housing Prices Dataset
https://www.kaggle.com/datasets/yasserh/housing-prices-dataset .

	Area (X1)	Bedrooms (X2)	Bathrooms (X3)	Stories (X4)	Price (Y)
1	7420	4	2	3	13300000
2	8960	4	4	4	12250000
3	9960	3	2	2	12250000
4	7500	4	2	2	12215000
5	7420	4	1	2	11410000
6	7500	3	3	1	10850000

7	8580	4	3	4	10150000
8	16200	5	3	2	10150000
9	8100	4	1	2	9870000
10	5750	3	2	4	9800000
11	13200	3	1	2	9800000
12	6000	4	3	2	9681000
13	6550	4	2	2	9310000
14	3500	4	2	2	9240000
15	7800	3	2	2	9240000
16	6000	4	1	2	9100000
17	6600	4	2	2	9100000
18	8500	3	2	4	8960000
19	4600	3	2	2	8890000
20	6420	3	2	2	8855000
21	4320	3	1	2	8750000
22	7155	3	2	1	8680000
23	8050	3	1	1	8645000
24	4560	3	2	2	8645000
25	8800	3	2	2	8575000
26	6540	4	2	2	8540000
27	6000	3	2	4	8463000
28	8875	3	1	1	8400000
29	7950	5	2	2	8400000
30	5500	4	2	2	8400000
31	7475	3	2	4	8400000
32	7000	3	1	4	8400000

33	4880	4	2	2	8295000
34	5960	3	3	2	8190000
35	6840	5	1	2	8120000
36	7000	3	2	4	8080940
37	7482	3	2	3	8043000
38	9000	4	2	4	7980000
39	6000	3	1	4	7962500
40	6000	4	2	4	7910000
41	6550	3	1	2	7875000
42	6360	3	2	4	7840000
43	6480	3	2	4	7700000
44	6000	4	2	4	7700000
45	6000	4	2	4	7560000
46	6000	3	2	3	7560000
47	6000	3	2	4	7525000
48	6600	3	1	4	7490000
49	4300	3	2	2	7455000
50	7440	3	2	1	7420000

3.1.2. Car Price Dataset

Car Price Dataset berasal dari Kaggle dan berisi 205 baris data dengan 26 kolom fitur terkait spesifikasi mobil. Fitur numerik utamanya meliputi panjang mobil (carlength), lebar mobil (carwidth), kapasitas mesin (enginesize), dan tenaga kuda (horsepower). Fitur lain seperti curb weight, wheelbase, fuelsystem dan tipe bodi bersifat kategorikal. Harga (price) menjadi variabel dependen yang ingin diprediksi. Untuk keperluan analisis manual, diambil 50 sampel data dengan empat fitur numerik tersebut sebagai variabel bebas. Tujuannya ialah mengukur pengaruh ukuran fisik dan performa mesin terhadap harga mobil.

Sumber Utama: Kaggle - Car Price Dataset

https://www.kaggle.com/datasets/ngawangchoeda/car-price-dataset .

	Car Lenght (X1)	Car Width (X2)	Engine Size (X3)	Horsepower (X4)	Price (Y)
1	168.8	64.1	130	111	13495
2	168.8	64.1	130	111	16500
3	171.2	65.5	152	154	16500
4	176.6	66.2	109	102	13950
5	176.6	66.4	136	115	17450
6	177.3	66.3	136	110	15250
7	192.7	71.4	136	110	17710
8	192.7	71.4	136	110	18920
9	192.7	71.4	131	140	23875
10	178.2	67.9	131	160	17859.2
11	176.8	64.8	108	101	16430
12	176.8	64.8	108	101	16925
13	176.8	64.8	164	121	20970
14	176.8	64.8	164	121	21105
15	189	66.9	164	121	24565
16	189	66.9	209	182	30760
17	193.8	67.9	209	182	41315
18	197	70.9	209	182	36880
19	141.1	60.3	61	48	5151
20	155.9	63.6	90	70	6295
21	158.8	63.6	90	70	6575
22	157.3	63.8	90	68	5572
23	157.3	63.8	90	68	6377

24	157.3	63.8	98	102	7957
25	157.3	63.8	90	68	6229
26	157.3	63.8	90	68	6692
27	157.3	63.8	90	68	7609
28	157.3	63.8	98	102	8558
29	174.6	64.6	122	88	8921
30	173.2	66.3	156	145	12964
31	144.6	63.9	92	58	6479
32	144.6	63.9	92	76	6855
33	150	64	79	60	5399
34	150	64	92	76	6529
35	150	64	92	76	7129
36	163.4	64	92	76	7295
37	157.1	63.9	92	76	7295
38	167.5	65.2	110	86	7895
39	167.5	65.2	110	86	9095
40	175.4	65.2	110	86	8845
41	175.4	62.5	110	86	10295
42	175.4	65.2	110	101	12945
43	169.1	66	110	100	10345
44	170.7	61.8	111	78	6785
45	155.9	63.6	90	70	8916.5
46	155.9	63.6	90	70	8916.5
47	172.6	65.2	119	90	11048
48	199.6	69.6	258	176	32250
49	199.6	69.6	258	176	35550

50	191.7	70.6	326	262	36000

3.1.3. Medical Cost Personal Dataset

Dataset ini berisi 1338 observasi yang mencatat informasi premi asuransi kesehatan individual. Setiap baris memuat variabel usia (age), jenis kelamin (sex), indeks massa tubuh (bmi), jumlah anak tanggungan (children), status perokok (smoker), wilayah tempat tinggal (region), serta biaya asuransi (charges). Dalam proyek ini diambil 50 observasi pertama. Variabel bebas yang dipilih adalah age, bmi, children, dan status smoker (1 = perokok, 0 = tidak), sedangkan charges menjadi variabel terikat. Analisis ini bertujuan mengidentifikasi seberapa besar faktor usia, kelebihan berat badan, jumlah anak, dan kebiasaan merokok mempengaruhi premi asuransi.

Sumber Utama: Kaggle - Medical Cost Personal Datasets

https://www.kaggle.com/datasets/mirichoi0218/insurance .

	Age (X1)	BMI (X2)	Children (X3)	Smoker (1), No Smoker (0) (X4)	Charges (Y)
1	19	27.9	0	1	13300000
2	18	33.77	1	0	12250000
3	28	33	3	0	12250000
4	33	22.705	0	0	12215000
5	32	28.88	0	0	11410000
6	31	25.74	0	0	10850000
7	46	33.44	1	0	10150000
8	37	27.74	3	0	10150000
9	37	29.83	2	0	9870000
10	60	25.84	0	0	9800000
11	25	26.22	0	0	9800000
12	62	26.29	0	1	9681000
13	23	34.4	0	0	9310000

14	56	39.82	0	0	9240000
15	27	42.13	0	1	9240000
16	19	24.6	1	0	9100000
17	52	30.78	1	0	9100000
18	23	23.845	0	0	8960000
19	56	40.3	0	0	8890000
20	30	35.3	0	1	8855000
21	60	36.005	0	0	8750000
22	30	32.4	1	0	8680000
23	18	34.1	0	0	8645000
24	34	31.92	1	1	8645000
25	37	28.025	2	0	8575000
26	59	27.72	3	0	8540000
27	63	23.085	0	0	8463000
28	55	32.775	2	0	8400000
29	23	17.385	1	0	8400000
30	31	36.3	2	1	8400000
31	22	35.6	0	1	8400000
32	18	26.315	0	0	8400000
33	19	28.6	5	0	8295000
34	63	28.31	0	0	8190000
35	28	36.4	1	1	8120000
36	19	20.425	0	0	8080940
37	62	32.965	3	0	8043000
38	26	20.8	0	0	7980000
39	35	36.67	1	1	7962500

40	60	39.9	0	1	7910000
41	24	26.6	0	0	7875000
42	31	36.63	2	0	7840000
43	41	21.78	1	0	7700000
44	37	30.8	2	0	7700000
45	38	37.05	1	0	7560000
46	55	37.3	0	0	7560000
47	18	38.665	2	0	7525000
48	28	34.77	0	0	7490000
49	60	24.53	0	0	7455000
50	36	35.2	1	1	7420000

3.1.4. Fish Market Dataset

Fish Market Dataset memuat ukuran fisik dan berat 50 ekor ikan dari berbagai spesies. Variabel yang direkam antara lain panjang vertikal (Length1), panjang diagonal (Length2), tinggi (Height), lebar (Width), dan berat ikan (Weight). Semua variabel bersifat numerik dengan satuan sentimeter untuk panjang/lebar/tinggi dan gram untuk berat. Variabel Weight digunakan sebagai variabel terikat, sedangkan empat variabel pengukuran badan ikan digunakan sebagai variabel bebas untuk memodelkan hubungan antara dimensi fisik dan massa ikan.

Sumber Utama: Kaggle - Fish Market Datasets

https://www.kaggle.com/datasets/abbasi1214/fish-market-dataset .

	Length1 (vertical length) (X1)	Length2 (diagonal length) (X2)	Height (X3)	Width (X4)	Weight (Y)
1	23.2	25.4	11.52	4.02	242
2	24	26.3	12.48	4.31	290
3	23.9	26.5	12.3778	4.70	340
4	26.3	29	12.73	4.46	363

5	26.5	29	12.444	5.13	430
6	26.8	29.7	13.6024	4.93	450
7	26.8	29.7	14.1795	5.28	500
8	27.6	30	12.67	4.69	390
9	27.6	30	14.0049	4.84	450
10	28.5	30.7	14.2266	4.96	500
11	28.4	31	14.2628	5.10	475
12	28.7	31	14.3714	4.81	500
13	29.1	31.5	13.7592	4.37	500
14	29.5	32	13.9129	5.07	340
15	29.4	32	14.9544	5.17	600
16	29.4	32	15.438	5.58	600
17	30.4	33	14.8604	5.29	700
18	30.4	33	14.938	5.20	700
19	30.9	33.5	15.633	5.13	610
20	31	33.5	14.4738	5.73	650
21	31.3	34	15.1285	5.57	575
22	31.4	34	15.9936	5.37	685
23	31.5	34.5	15.5227	5.28	620
24	31.8	35	15.4686	6.13	680
25	31.9	35	16.2405	5.59	700
26	31.8	35	16.36	6.05	725
27	32	35	16.3618	6.09	720
28	32.7	36	16.517	5.85	714
29	32.8	36	16.8896	6.20	850
30	33.5	37	18.957	6.60	1000

31	35	38.5	18.0369	6.31	920
32	35	38.5	18.084	6.29	955
33	36.2	39.5	18.7542	6.75	925
34	37.4	41	18.6354	6.75	975
35	38	41	17.6235	6.37	950
36	12.9	14.1	4.1472	2.27	40
37	16.5	18.2	5.2983	2.82	69
38	17.5	18.8	5.5756	2.90	78
39	18.2	19.8	5.6166	3.17	87
40	18.6	20	6.216	3.57	120
41	22	23.4	6.9153	3.63	161
42	19.1	20.8	6.1677	3.40	110
43	19.4	21	6.1146	3.29	120
44	20.4	22	5.8045	3.75	150
45	20.5	22	6.6339	3.55	145
46	20.5	22.5	7.0334	3.82	160
47	21	22.5	6.55	3.33	140
48	21.1	22.5	6.4	3.80	160
49	22	24	7.5344	3.84	169
50	22	23.4	6.9153	3.63	161

3.1.5. 3D Printer Dataset

Dataset ini berasal dari eksperimen percetakan 3D dan mencakup sembilan parameter pengaturan antara lain layer height, wall thickness, infill density, infill pattern, nozzle temperature, bed temperature, print speed, material, dan fan speed serta tiga variabel keluaran: roughness, tension strength, dan elongation. Untuk analisis regresi ini diambil 50 sampel dan dipilih empat variabel bebas: tinggi lapisan (layer_height),

kepadatan isi (infill_density), temperatur nozzle (nozzle_temperature), serta kecepatan cetak (print_speed). Variabel terikat yang diprediksi adalah tension_strength, yaitu kekuatan tarik hasil cetak. Dataset ini berguna untuk memahami pengaruh pengaturan printer terhadap sifat mekanik objek cetak.

Sumber Utama: Kaggle - 3D Printer Dataset

https://www.kaggle.com/datasets/afumetto/3dprinter .

	layer_height (X1)	infill_density (X2)	nozzle_temperature (X3)	print_speed (X4)	tension_strenght (Y)
1	0.02	90	220	40	18
2	0.02	90	225	40	16
3	0.02	80	230	40	8
4	0.02	70	240	40	10
5	0.02	90	250	40	5
6	0.02	40	200	40	24
7	0.02	10	205	40	12
8	0.02	10	210	40	14
9	0.02	70	215	40	27
10	0.02	40	220	40	25
11	0.06	80	220	60	37
12	0.06	20	225	60	12
13	0.06	50	230	60	16
14	0.06	10	240	60	9
15	0.06	50	250	60	10
16	0.06	90	200	60	27
17	0.06	40	205	60	23
18	0.06	30	210	60	26
19	0.06	80	215	60	33
20	0.06	50	220	60	29

21	0.1	40	220	120	16
22	0.1	30	225	120	12
23	0.1	50	230	120	10
24	0.1	80	240	120	19
25	0.1	60	250	120	8
26	0.1	50	200	120	11
27	0.1	40	205	120	12
28	0.1	50	210	120	18
29	0.1	90	215	120	34
30	0.1	30	220	120	14
31	0.15	50	220	60	27
32	0.15	10	225	60	19
33	0.15	50	230	60	18
34	0.15	50	240	60	9
35	0.15	80	250	60	13
36	0.15	80	200	60	33
37	0.15	50	205	60	24
38	0.15	30	210	60	26
39	0.15	40	215	60	22
40	0.15	10	220	60	4
41	0.2	80	220	40	35
42	0.2	90	225	40	34
43	0.2	30	230	40	28
44	0.2	90	240	40	28
45	0.2	80	250	40	21
46	0.2	60	200	40	28

47	0.2	20	205	40	14
48	0.2	60	210	40	30
49	0.2	40	215	40	29
50	0.2	60	220	40	27

3.2. Implementasi Metode Multiple Regresi Linier Terhadap Housing Price Dataset

3.2.1. Tahap Pengerjaan

Membentuk matriks desain X dengan ukuran $m \times (n+1)$, di mana m adalah jumlah observasi dan n jumlah variabel bebas. Kolom pertama diisi dengan angka 1 untuk merepresentasikan intersep, sedangkan kolom berikutnya diisi variabel bebas (X1–X4). Menyusun vektor target Y dengan panjang m yang memuat variabel terikat pada setiap observasi.

Sigma (X1)	Sigma (X2)	Sigma (X3)	Sigma (X4)	Sigma (Y)	Sigma X1^2
353677	174	95	132.00	447425440	2716849699

Sigma (X2^2)	Sigma (X3^3)	Sigma (X4^2)
624	201	404

Sigma (X1X2)	Sigma (X1X3)	Sigma (X1X4)	Sigma (X2X3)	Sigma (X2X4)	Sigma (X3X4)
1243561	680559	924646	334	457	255

Sigma (X1Y)	Sigma (X2Y)	Sigma (X3Y)	Sigma (X4Y)
3216186806000	1567627320	862549380	1168101760

3.2.2. Menyusun Matriks Augmented

$$\begin{bmatrix} N & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum (X1^2) & \sum (X2.X1) & \sum (X3.X1) & \sum (X4.X1) \\ \sum X2 & \sum (X1.X2) & \sum (X2^2) & \sum (X3.X2) & \sum (X4.X2) \\ \sum X3 & \sum (X1X3) & \sum (X2.X3) & \sum (X3^2) & \sum (X4.X3) \\ \sum X4 & \sum (X1.X4) & \sum (X2.X4) & \sum (X3.X4) & \sum (X4^2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \sum (Y) \\ \sum (X1.Y) \\ \sum (X2.Y) \\ \sum (X3.Y) \\ \sum (X4.\ Y) \end{bmatrix}$$

Membentuk normal equation dengan menghitung $A = X^T X$ dan $B = X^T Y$. Matriks A berukuran $(n + 1) \times (n + 1)$ dan vektor B berukuran $(n + 1) \times 1$.

Matriks Augmented					
50	353677	174	95	132.00	447425440
353677	2716849699	1243561	680559	924646	3216186806000
174	1243561	624	334	457	1567627320
95	680559	334	201	255	862549380
132.00	924646	457	255	404	1168101760

3.2.3. Tranformasi ke Bentuk Segitiga Atas

Eliminasi Gauss: Matriks augmented $[AB]$ direduksi menjadi bentuk segitiga atas menggunakan eliminasi baris elementer. Proses pivoting dilakukan untuk menghindari pembagian oleh angka mendekati nol. Penulis membagi seluruh baris pertama dengan 50 supaya elemen pertama menjadi 1.

$$R_1 \leftarrow \frac{1}{50} R_1$$

Matriks Bentuk Segitiga Atas Tahap 1 (R1 <- 1/50R1)					
1	7073.54	3.48	1.9	2.64	8948508.8
353677	2716849699	1243561	680559	924646	3.21619E+12

174	1243561	624	334	457	1567627320
95	680559	334	201	255	862549380
132.00	924646	457	255	404	1168101760

Penulis menginginkan kolom pertama menjadi seperti:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Jadi penulis melakukan:

$$R_2 \leftarrow R_2 - 353677R_1$$

$$R_3 \leftarrow R_3 - 174R_1$$

$$R_4 \leftarrow R_4 - 95R_1$$

$$R_5 \leftarrow R_5 - 132R_1$$

Setelah eliminasi bentuknya menjadi:

Matriks Bentuk Segitiga Atas Tahap (R2 <-R2 - 353677R1, R3 <-R3 - 174R1, R4 <-R4 - 95R1, R5 <-R5 - 132R1)					
1	7073.54	3.48	1.9	2.64	8948508.8
0	215101292.4	12765.04	8572.7	-9061.28	51305059142
0	12765.04	18.48	3.4	-2.36	10586788.8
0	8572.7	3.4	20.5	4.2	12441044
0	-9061.28	-2.36	4.2	55.52	-13101401.6

Matriks Bentuk Segitiga Atas Tahap ($R_3 \leftarrow R_3 - 12765.04R_2$, $R_4 \leftarrow R_4 - 8572.7R_2$, $R_5 \leftarrow R_5$ - (-9061.28)R_2)					
1	7073.54	3.48	1.9	2.64	8948508.8
0	1	0.0001	0.0000	0.0000	238.5158

0	0	17.7225	2.8913	-1.8223	7542124.9359
0	0	2.8913	20.1583	4.5611	10396319.5097
0	0	-1.8223	4.5611	55.1383	-10940143.0549

Matriks Bentuk Segitiga Atas Tahap (Normalisasi)					
1	7073.54	3.48	1.9	2.64	8948508.8
0	1	0.0001	0.0000	0.0000	238.5158
0	0	1	0.1631	-0.1028	425568.5608
0	0	0	19.6867	4.8584	9165890.6014
0	0	0	4.8584	54.9509	-10164644.5754

Matriks Bentuk Segitiga Atas Tahap (Normalisasi R4)					
1	7073.54	3.48	1.9	2.64	8948508.8
0	1	0.0001	0.0000	0.0000	238.5158
0	0	0	0.0092	-0.0058	24012.9408
0	0	0	1	0.2468	465588.9360
0	0	0	0	53.7519	-12426669.4631

Setelah matriks berada pada bentuk segitiga atas, koefisien regresi diperoleh b_0, b_1, b_2, b_3, b_4 dengan menyelesaikan sistem persamaan linear dari baris paling bawah ke atas.

Substitusi Mundur	
b0	6126230.3808

b1	189.1630
b2	316533.1602
b3	522642.5880
b4	-231185.5786

	Y Prediksi	e = Y Aktual - Y Prediksi
1	9147680.9061	4152319.0939
2	10253091.5202	1996908.4798
3	9542807.3390	2707192.6610
4	9393999.5245	2821000.4755
5	8856223.8967	2553776.1033
6	9831294.5308	1018705.4692
7	9658566.9930	491433.0070
8	11878893.3541	-1728893.3541
9	8984854.7352	885145.2648
10	8284059.9608	1515940.0392
11	9633052.8640	166947.1360
12	9632897.6157	48102.3843
13	9214294.6765	95705.3235
14	8637347.5330	602652.4670
15	9134215.2636	105784.7364
16	8587612.4397	512387.5603
17	9223752.8264	-123752.8264
18	8804258.2050	155741.7950
19	8528893.6704	361106.3296
20	8873170.3265	-18170.3265

21	7953285.4431	796714.5569
22	9243390.7085	-563390.7085
23	8890049.0036	-245049.0036
24	8521327.1505	123672.8495
25	9323378.2614	-748378.2614
26	9212403.0465	-672403.0465
27	8331350.7103	131649.2897
28	9046108.4769	-646108.4769
29	9795656.0338	-1395656.0338
30	9015673.5288	-615673.5288
31	8610366.1321	-210366.1321
32	7997871.1202	402128.8798
33	8898392.4701	-603392.4701
34	9308797.9355	-1118797.9355
35	9063042.5182	-943042.5182
36	8520513.7082	-439573.7082
37	8842875.8517	-799875.8517
38	9215372.8641	-1235372.8641
39	7808708.1223	153791.8777
40	8647883.8706	-737883.8706
41	8375118.9283	-500118.9283
42	8399449.3895	-559449.3895
43	8422148.9493	-722148.9493
44	8647883.8706	-947883.8706
45	8647883.8706	-1087883.8706
46	8562536.2889	-1002536.2889

47	8331350.7103	-806350.7103
48	7922205.9211	-432205.9211
49	8472144.7711	-1017144.7711
50	9297302.1629	-1877302.1629

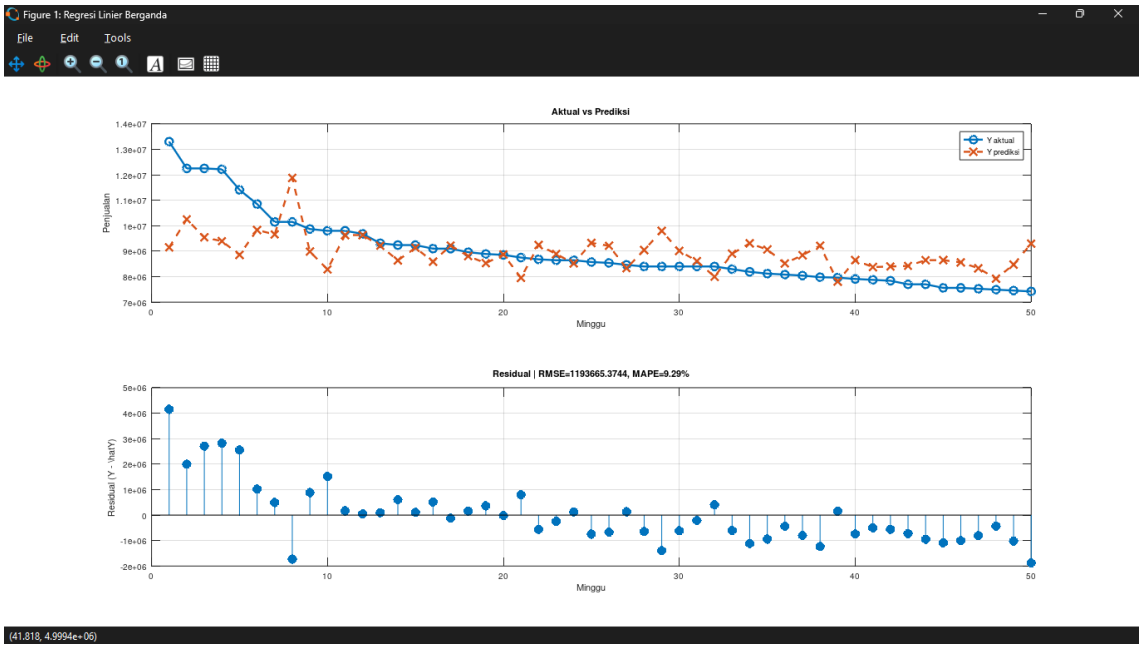
3.2.4. Nilai Error

Vektor koefisien b digunakan untuk menghitung prediksi dengan rumus $\hat{Y} = Xb$,. Selisih antara nilai aktual dan prediksi Y - $\hat{Y}$ membentuk residual yang dianalisis untuk mengidentifikasi outlier.

RSS	71241851299203.0000
MSE	1424837025984.0600
RMSE	1193665.3744
MAPE	9.2913

Menurut data dictionary housing.csv di RPubs, variabel price menunjukkan harga rumah dalam Rupee (Rs.), sedangkan area adalah ukuran properti dalam kaki persegi. Kolom bedrooms, bathrooms, dan stories berupa jumlah ruangan sehingga tidak memiliki satuan unit. Hasil regresi menunjukkan bahwa luas rumah memiliki pengaruh terbesar terhadap harga. Setiap kenaikan 1 kaki persegi menambah harga sekitar 189 rupiah. Penambahan satu kamar tidur meningkatkan harga rata-rata 316 533 rupiah, sedangkan satu kamar mandi menambah sekitar 522.643 rupiah. Uniknya, koefisien untuk stories negatif (~231.186 rupiah), yang berarti rumah bertingkat lebih dari satu cenderung dihargai lebih rendah. Kemungkinan karena rumah bertingkat di wilayah tersebut memiliki luas lantai yang lebih kecil. Secara keseluruhan, variabel area mendominasi model, sedangkan jumlah kamar tidur dan kamar mandi memiliki efek moderat.

3.2.5. Outlier



Pada analisis regresi linier berganda, outlier adalah data yang memiliki nilai residual jauh lebih besar atau lebih kecil dibandingkan observasi lainnya. Kehadiran outlier dapat mengganggu kemiringan garis regresi dan menurunkan kualitas model, sehingga identifikasi dan penanganannya menjadi penting.

Gambar menunjukkan plot Aktual vs Prediksi dan grafik residual untuk Housing Prices Dataset. Pada grafik pertama, garis biru merepresentasikan harga rumah aktual, sedangkan tanda silang oranye adalah hasil prediksi dari model. Secara umum, prediksi mengikuti tren menurun harga rumah terhadap minggu, namun terdapat beberapa titik prediksi (sekitar minggu ke-8 sampai ke-15) yang menyimpang cukup jauh dari nilai aktual.

Grafik residual di bawahnya memvisualisasikan selisih antara nilai aktual dan nilai prediksi ($Y - \hat{Y}$) untuk setiap minggu. Titik residual yang jauh di atas atau di bawah garis nol menandakan adanya outlier dalam data. Misalnya, pada awal minggu terlihat residual positif yang sangat tinggi, yang menandakan bahwa model meremehkan harga sebenarnya pada periode tersebut. Sebaliknya, residual negatif ekstrem menunjukkan prediksi yang terlalu tinggi dibanding harga aktual. Dengan menganalisis pola residual

ini, mahasiswa dapat memutuskan apakah perlu melakukan transformasi variabel, menghapus data yang tidak wajar, atau memilih model yang lebih kompleks agar outlier tidak mendistorsi hasil estimasi.

3.2.6. *Source Code*

```
%====== Multiple Regresi Linier ======
A = [50,      353677,        174,        95,
132;
     353677,  2716849699,  1243561,   680559,
924646;
     174,     1243561,        624,        334,
457;
     95,      680559,         334,        201,
255;
     132,     924646,         457,        255,
404];

B = [447425440;
     3216186806000;
     1567627320;
     862549380;
     1168101760];

AB = [A, B];
[n, ~] = size(AB);
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
```

```
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);

X1 = [7420; 8960; 9960; 7500; 7420; 7500; 8580;
16200; 8100; 5750;
      13200; 6000; 6550; 3500; 7800; 6000; 6600;
8500; 4600; 6420;
      4320; 7155; 8050; 4560; 8800; 6540; 6000;
8875; 7950; 5500;
      7475; 7000; 4880; 5960; 6840; 7000; 7482;
9000; 6000; 6000;
      6550; 6360; 6480; 6000; 6000; 6000; 6000;
6600; 4300; 7440];
```

```
X2 = [4; 4; 3; 4; 4; 3; 4; 5; 4; 3;
      3; 4; 4; 4; 3; 4; 4; 3; 3; 3;
      3; 3; 3; 3; 3; 4; 3; 3; 5; 4;
      3; 3; 4; 3; 5; 3; 3; 4; 3; 4;
      3; 3; 3; 4; 4; 3; 3; 3; 3; 3];

X3 = [2; 4; 2; 2; 1; 3; 3; 3; 1; 2;
      1; 3; 2; 2; 2; 1; 2; 2; 2; 2;
      1; 2; 1; 2; 2; 2; 2; 1; 2; 2;
      2; 1; 2; 3; 1; 2; 2; 2; 1; 2;
      1; 2; 2; 2; 2; 2; 2; 1; 2; 2];

X4 = [3; 4; 2; 2; 2; 1; 4; 2; 2; 4;
      2; 2; 2; 2; 2; 2; 2; 4; 2; 2;
      2; 1; 1; 2; 2; 2; 4; 1; 2; 2;
      4; 4; 2; 2; 2; 4; 3; 4; 4; 4;
      2; 4; 4; 4; 4; 3; 4; 4; 2; 1];

Y  = [13300000; 12250000; 12250000; 12215000;
11410000;
      10850000; 10150000; 10150000; 9870000;
9800000;
      9800000; 9681000; 9310000; 9240000; 9240000;
9100000;
      9100000; 8960000; 8890000; 8855000; 8750000;
8680000;
      8645000; 8645000; 8575000; 8540000; 8463000;
8400000;
```

```
      8400000; 8400000; 8400000; 8400000; 8295000;
8190000;
      8120000; 8080940; 8043000; 7980000; 7962500;
7910000;
      7875000; 7840000; 7700000; 7700000; 7560000;
7560000;
      7525000; 7490000; 7455000; 7420000];

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];

Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);

% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
```

```
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

3.2.7. Penjelasan *Source Code*

1. Menyusun sistem normal equations (A, B)

```
%====== Multiple Regresi Linier ======
A = [50, 353677, 174, 95, 132;
     353677,  2716849699, 1243561, 680559, 924646;
     174, 1243561, 624, 334, 457;
     95, 680559, 334, 201, 255;
     132, 924646, 457, 255, 404];

B = [447425440;
3216186806000;
1567627320;
862549380;
1168101760];

AB = [A, B];
[n, ~] = size(AB);
```

Bagian ini memuat matriks normal equations hasil $X^T X$ (disimpan sebagai A) dan vektor $X^T y$ (disimpan sebagai B). Keduanya digabung menjadi matriks teraugmentasi AB = [A B] berukuran 5×6. Nilai n = 5 adalah banyaknya koefisien yang ingin dicari: b_0, b_1, b_2, b_3, b_4.

2. Eliminasi Gauss dengan partial pivoting

```
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
```

Partial pivoting: memilih elemen terbesar (secara absolut) pada kolom ke-i di baris i..n sebagai pivot, lalu menukar baris bila perlu. Tujuannya menstabilkan perhitungan dan mengurangi pembagian dengan nilai kecil.

Eliminasi maju: untuk setiap baris di bawah pivot (indeks j), dihitung faktor Factor = a_{ji}/a_{ii} kemudian dikurangi baris j dengan factor × baris i. Hasilnya, elemen di bawah pivot menjadi nol dan terbentuk matriks segitiga atas.

3. Substitusi balik → mendapatkan koefisien *b*

```
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
```

```
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);
```

Setelah bentuk segitiga atas diperoleh, dilakukan back-substitution: untuk baris i dihitung jumlah kontribusi kolom kanan $s = \sum_{kk=i+1}^{n} a_{ik} b_{k`}$ lalu $b_i = \frac{a_{i,n+1} - s}{a_{ii}}$. Hasil akhirnya adalah vektor koefisien regresi $b = [b_0, \ldots, b_4]^T$.

4. Menyusun data asli dan matriks desain *X*

```
X1 = [7420; 8960; 9960; 7500; 7420; 7500; 8580;
16200; 8100; 5750;
      13200; 6000; 6550; 3500; 7800; 6000; 6600;
8500; 4600; 6420;
      4320; 7155; 8050; 4560; 8800; 6540; 6000;
8875; 7950; 5500;
      7475; 7000; 4880; 5960; 6840; 7000; 7482;
9000; 6000; 6000;
      6550; 6360; 6480; 6000; 6000; 6000; 6000;
6600; 4300; 7440];

X2 = [4; 4; 3; 4; 4; 3; 4; 5; 4; 3;
      3; 4; 4; 4; 3; 4; 4; 3; 3; 3;
      3; 3; 3; 3; 3; 4; 3; 3; 5; 4;
      3; 3; 4; 3; 5; 3; 3; 4; 3; 4;
      3; 3; 3; 4; 4; 3; 3; 3; 3; 3];

X3 = [2; 4; 2; 2; 1; 3; 3; 3; 1; 2;
```

```
      1; 3; 2; 2; 2; 1; 2; 2; 2; 2;
      1; 2; 1; 2; 2; 2; 2; 1; 2; 2;
      2; 1; 2; 3; 1; 2; 2; 2; 1; 2;
      1; 2; 2; 2; 2; 2; 2; 1; 2; 2];

X4 = [3; 4; 2; 2; 2; 1; 4; 2; 2; 4;
      2; 2; 2; 2; 2; 2; 2; 4; 2; 2;
      2; 1; 1; 2; 2; 2; 4; 1; 2; 2;
      4; 4; 2; 2; 2; 4; 3; 4; 4; 4;
      2; 4; 4; 4; 4; 3; 4; 4; 2; 1];

Y  = [13300000; 12250000; 12250000; 12215000;
11410000;
      10850000; 10150000; 10150000; 9870000;
9800000;
      9800000; 9681000; 9310000; 9240000; 9240000;
9100000;
      9100000; 8960000; 8890000; 8855000; 8750000;
8680000;
      8645000; 8645000; 8575000; 8540000; 8463000;
8400000;
      8400000; 8400000; 8400000; 8400000; 8295000;
8190000;
      8120000; 8080940; 8043000; 7980000; 7962500;
7910000;
      7875000; 7840000; 7700000; 7700000; 7560000;
7560000;
      7525000; 7490000; 7455000; 7420000];
```

```
m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];
```

Pada potongan kode ini, kita mendefinisikan empat variabel bebas (X1–X4), variabel terikat (Y), serta menyiapkan matriks X untuk proses regresi. Masing-masing variabel bebas menyimpan 50 nilai sesuai jumlah observasi pada Housing Prices Dataset, yaitu:

- X1 merupakan vektor yang merepresentasikan Area rumah (luas bangunan) untuk setiap observasi. Nilai-nilainya dituliskan secara vertikal menggunakan tanda titik koma ; agar terbentuk vektor kolom di Octave/MATLAB.
- X2 berisi jumlah kamar tidur (Bedrooms) untuk masing-masing rumah. Nilai 3, 4, dan 5 menunjukkan banyaknya kamar tidur dalam satu unit rumah.
- X3 adalah jumlah kamar mandi (Bathrooms) per rumah. Nilai 1–3 merefleksikan berapa kamar mandi yang dimiliki oleh setiap rumah.
- X4 menyimpan data jumlah lantai (Stories) dari tiap rumah. Angka 1–4 menandakan berapa tingkat bangunan pada rumah tersebut.
- Y adalah vektor target yang merepresentasikan harga jual (Price) setiap rumah dalam satuan rupiah. Vektor ini memuat 50 nilai, masing-masing sesuai dengan urutan baris pada X1–X4.

Setelah semua vektor dideklarasikan, kode berikut menghitung ukuran sampel:

```
m  = rows(Y);
```

Fungsi rows(Y) mengembalikan jumlah baris dari vektor Y, yang dalam konteks ini sama dengan jumlah observasi (50). Nilai m akan digunakan untuk membangun matriks desain dan menghitung metrik error.

Matriks desain X kemudian dibentuk dengan menambahkan satu kolom konstanta (intersep) di bagian awal dan diikuti oleh keempat variabel bebas: `X = [ones(m,1), X1, X2, X3, X4];`

Perintah ones(m,1) menghasilkan sebuah vektor kolom berukuran m×1 yang seluruh elemennya bernilai 1. Kolom ini diperlukan untuk menghitung koefisien

intersep b0 dalam model regresi. Selanjutnya, X1, X2, X3, dan X4 disusun berdampingan untuk membentuk matriks X berukuran m×5, di mana setiap baris mewakili satu observasi dan setiap kolom (setelah kolom pertama) mewakili satu variabel bebas.

5. Prediksi dan metrik galat

```
Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);
```

$\hat{Y} = Xb$ adalah prediksi, Residual e = Y - $\hat{Y}$, MSE = SSE dibagi jumlah observasi, RMSE = akar dari MSE → satuannya sama dengan Y, MAPE menyatakan rata-rata persentase kesalahan absolut (dalam %). Penulisan %%%% pada sprintf/printf memang diperlukan untuk mencetak tanda % di output.

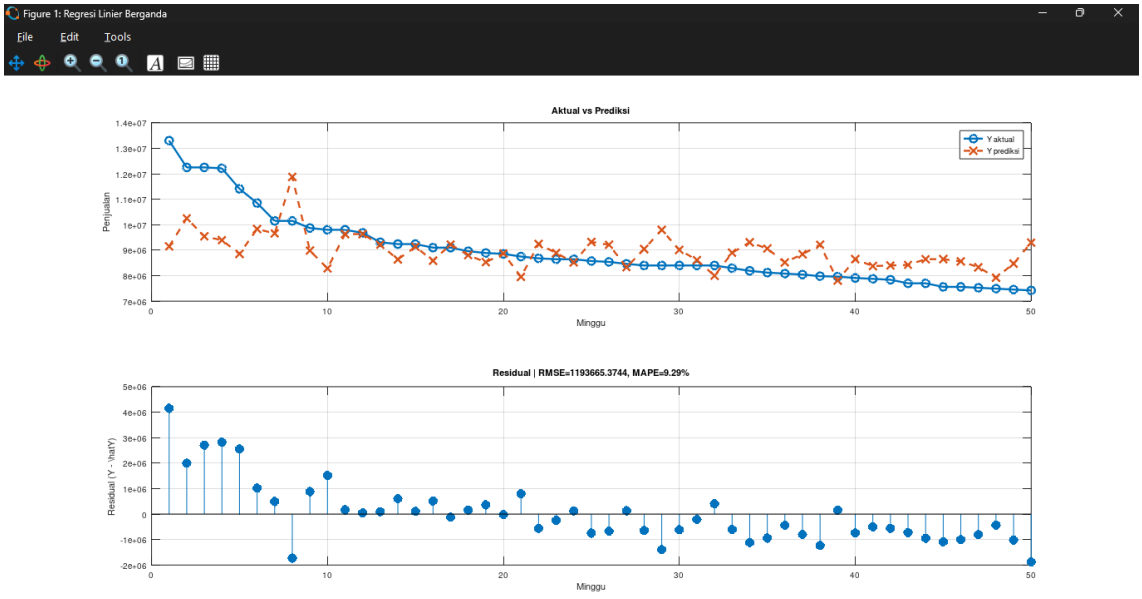
6. Grafik

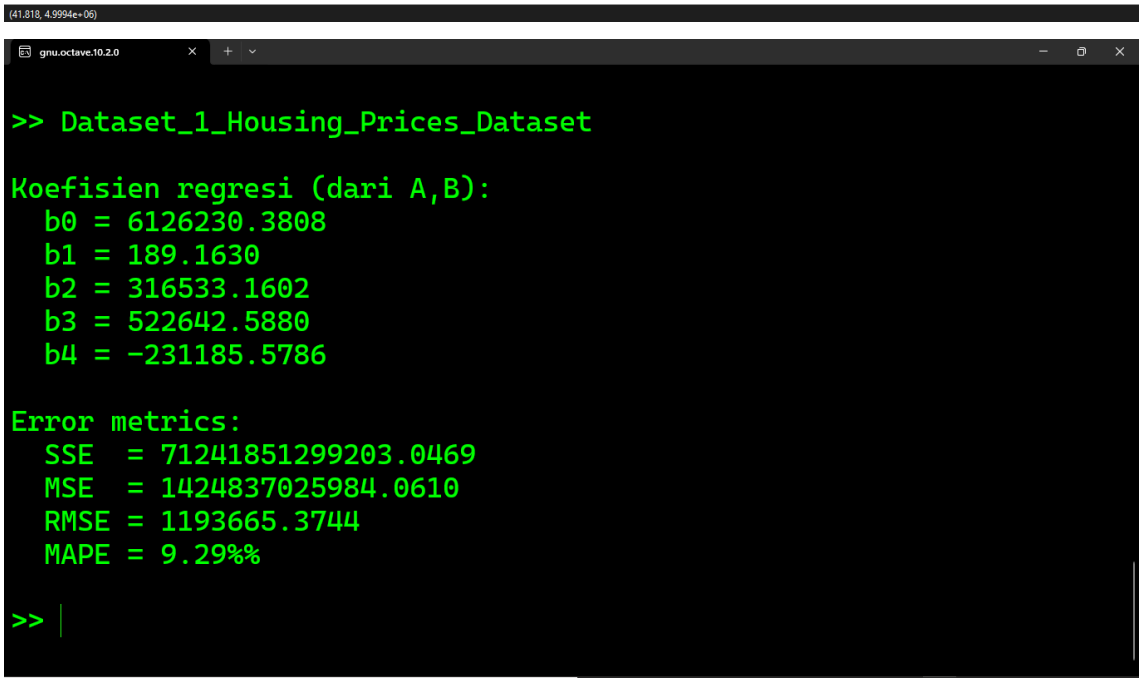
```
% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
```

```
legend('Y                                    aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

- Subplot atas: membandingkan kurva aktual dan prediksi untuk melihat kedekatan pola.
- Subplot bawah: menampilkan residual per minggu. Titik yang jauh dari garis nol dapat diinterpretasi sebagai kandidat outlier atau observasi yang kurang terjelaskan oleh model.
  Di title(sprintf(...)) tutup format MAPE dengan %% agar tanda persen tercetak.

3.2.8. Tampilan Grafik dan Terminal





3.3. Implementasi Metode Multiple Regresi Linier Terhadap Car Price Dataset

3.3.1. Tahap Pengerjaan

Membentuk matriks desain X dengan ukuran $m \times (n+1)$, di mana m adalah jumlah observasi dan n jumlah variabel bebas. Kolom pertama diisi dengan angka 1 untuk merepresentasikan intersep, sedangkan kolom berikutnya diisi variabel bebas (X1–X4).

Menyusun vektor target Y dengan panjang m yang memuat variabel terikat pada setiap observasi.

Sigma (X1)	Sigma (X2)	Sigma (X3)	Sigma (X4)	Sigma (Y)	Sigma X1^2
8512.3	3272.5	6370	5264.00	715226.167	1460891.29

Sigma (X2^2)	Sigma (X3^3)	Sigma (X4^2)
214504.71	942710	642194

Sigma (X1X2)	Sigma (X1X3)	Sigma (X1X4)	Sigma (X2X3)	Sigma (X2X4)	Sigma (X3X4)
558743.89	1115424.5	921589	421508.5	348495.6	771326

Sigma (X1Y)	Sigma (X2Y)	Sigma (X3Y)	Sigma (X4Y)
127932425	47723006.24	112941169.9	93128379.72

3.3.2. Menyusun Matriks Augmented

$$\begin{bmatrix} N & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum (X1^2) & \sum (X2.X1) & \sum (X3.X1) & \sum (X4.X1) \\ \sum X2 & \sum (X1.X2) & \sum (X2^2) & \sum (X3.X2) & \sum (X4.X2) \\ \sum X3 & \sum (X1X3) & \sum (X2.X3) & \sum (X3^2) & \sum (X4.X3) \\ \sum X4 & \sum (X1.X4) & \sum (X2.X4) & \sum (X3.X4) & \sum (X4^2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \sum (Y) \\ \sum (X1.Y) \\ \sum (X2.Y) \\ \sum (X3.Y) \\ \sum (X4.\ Y) \end{bmatrix}$$

Membentuk normal equation dengan menghitung $A = X^T X$ dan $B = X^T Y$. Matriks A berukuran $(n+1) \times (n+1)$ dan vektor B berukuran $(n+1) \times 1$.

Matriks Augmented					
50	8512.3	3272.5	6370	5264.00	715226.1670
8512.3	1460891.29	558743.89	1115424.5	921589	127932424.6594
3272.5	558743.89	214504.71	421508.5	348495.6	47723006.2393

6370	1115424.5	421508.5	942710	771326	112941169.8770
5264.00	921589	348495.6	771326	642194	93128379.7200

3.3.3. Tranformasi ke Bentuk Segitiga Atas

Eliminasi Gauss: Matriks augmented $[AB]$ direduksi menjadi bentuk segitiga atas menggunakan eliminasi baris elementer. Proses pivoting dilakukan untuk menghindari pembagian oleh angka mendekati nol. Penulis membagi seluruh baris pertama dengan 50 supaya elemen pertama menjadi 1.

$$R_1 \leftarrow \frac{1}{50} R_1$$

Matriks Bentuk Segitiga Atas Tahap 1 (R1 <- 1/50R1)					
1	170.246	65.45	127.4	105.28	14304.52334
8512.3	1460891.29	558743.89	1115424.5	921589	127932424.7
3272.5	558743.89	214504.71	421508.5	348495.6	47723006.24
6370	1115424.5	421508.5	942710	771326	112941169.9
5264.00	921589	348495.6	771326	642194	93128379.72

Penulis menginginkan kolom pertama menjadi seperti:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Jadi penulis melakukan:

$$R_2 \leftarrow R_2 - 8512.3R_1$$
$$R_3 \leftarrow R_3 - 3272.5R_1$$
$$R_4 \leftarrow R_4 - 6370R_1$$
$$R_5 \leftarrow R_5 - 5264R_1$$

Setelah eliminasi bentuknya menjadi:

Matriks Bentuk Segitiga Atas Tahap (R2 <-R2 - 8512.3R1, R3 <-R3 - 3272.5R1, R4 <-R4 - 6370R1, R5 <-R5 - 5264R1)

1	170.246	65.45	127.4	105.28	14304.52334
0	11706.2642	1613.855	30957.48	25414.056	6168030.632
0	1613.855	319.585	4592	3966.8	911453.6092
0	30957.48	4592	131172	100692.4	21821356.2
0	25414.056	3966.8	100692.4	88000.08	17829368.86

Matriks Bentuk Segitiga Atas Tahap ($R_3 \leftarrow R_3 - 1613.855R_2$, $R_4 \leftarrow R_4 - 30957.48R_2$, $R_5 \leftarrow R_5 - 25414.056R_2$)					
1	170.246	65.45	127.4	105.28	14304.52334
0	1	0.1379	2.6445	2.1710	526.9000
0	0	97.0949	324.1240	463.1544	61113.4062
0	0	324.1240	49304.2451	33484.3551	5509859.9222
0	0	463.1544	33484.3551	32826.6932	4438702.6968

Matriks Bentuk Segitiga Atas Tahap (Normalisasi)					
1	170.246	65.45	127.4	105.28	14304.52334
0	1	0.1379	2.6445	2.1710	526.9000
0	0	1	3.3382	4.7701	629.4194
0	0	0	48222.2482	31938.2444	5305849.9938
0	0	0	31938.2444	30617.3909	4147184.3662

Matriks Bentuk Segitiga Atas Tahap (Normalisasi R4)					
1	170.246	65.45	127.4	105.28	14304.52334
0	1	0.1379	2.6445	2.1710	526.9000
0	0	0	0.0344	0.0491	6.4825
0	0	0	1	0.6623	110.0291

0	0	0	0	9464.2616	633048.4654

Setelah matriks berada pada bentuk segitiga atas, koefisien regresi diperoleh b_0, b_1, b_2, b_3, b_4 dengan menyelesaikan sistem persamaan linear dari baris paling bawah ke atas.

Substitusi Mundur	
b0	-40317.4354
b1	195.3304
b2	90.9394
b3	65.7281
b4	66.8883

	Y Prediksi	e = Y - Y Prediksi
1	14452.8015	-957.8015
2	14452.8015	2047.1985
3	19371.1243	-2871.1243
4	14185.0673	-235.0673
5	16847.4608	602.5392
6	16640.6566	-1390.6566
7	20112.5357	-2402.5357
8	20112.5357	-1192.5357
9	21790.5447	2084.4553
10	19977.7322	-2118.5652
11	13964.2019	2465.7981
12	13964.2019	2960.7981
13	18982.7394	1987.2606

14	18982.7394	2122.2606
15	21556.7430	3008.2570
16	28594.6926	2165.3074
17	29623.2180	11691.7820
18	30521.0934	6358.9066
19	-52.6200	5203.6200
20	6516.0265	-221.0265
21	7082.4847	-507.4847
22	6673.9003	-1101.9003
23	6673.9003	-296.9003
24	9473.9274	-1516.9274
25	6673.9003	-444.9003
26	6673.9003	18.0997
27	6673.9003	935.0997
28	9473.9274	-915.9274
29	13566.9319	-4645.9319
30	19495.4540	-6531.4540
31	3664.8712	2814.1288
32	4868.8608	1986.1392
33	4008.0611	1390.9389
34	5932.7389	596.2611
35	5932.7389	1196.2611
36	8550.1663	-1255.1663
37	7310.4908	-15.4908
38	11312.1364	-3417.1364
39	11312.1364	-2217.1364

40	12855.2465	-4010.2465
41	12609.7102	-2314.7102
42	13858.5712	-913.5712
43	12633.8529	-2288.8529
44	11158.6213	-4373.6213
45	6516.0265	2400.4735
46	6516.0265	2400.4735
47	13167.4272	-2119.4272
48	33730.0763	-1480.0763
49	33730.0763	1819.9237
50	42499.8082	-6499.8082

3.3.4. Nilai Error

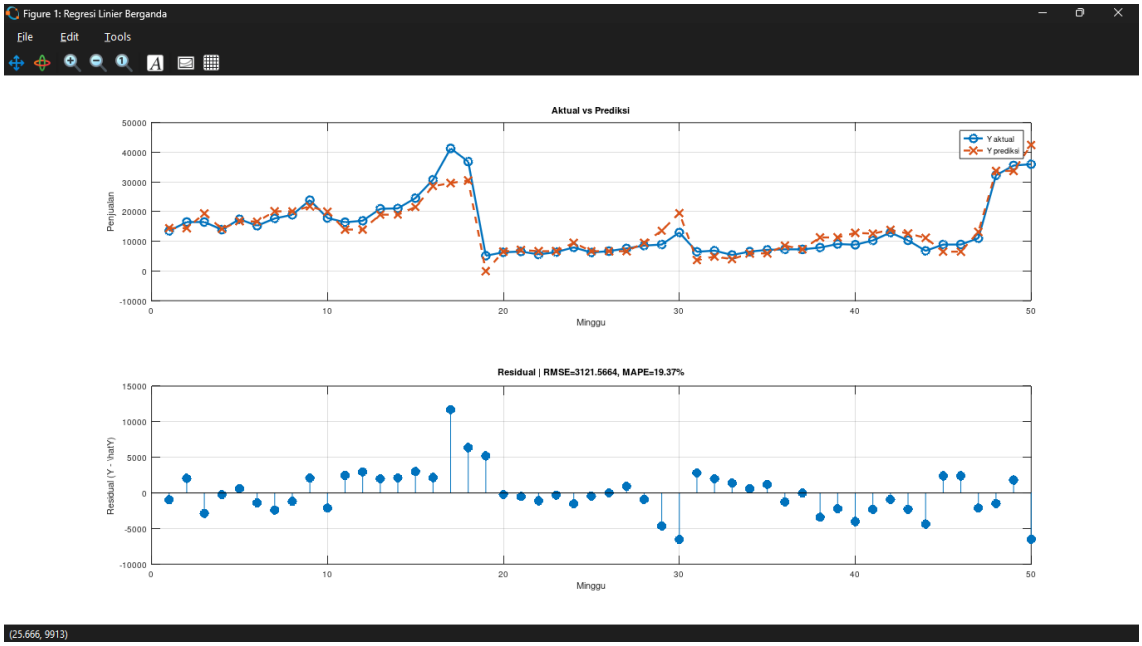
Vektor koefisien b digunakan untuk menghitung prediksi dengan rumus $\hat{Y} = Xb$,. Selisih antara nilai aktual dan prediksi Y - $\hat{Y}$ membentuk residual yang dianalisis untuk mengidentifikasi outlier.

RSS	487208826.4937
MSE	9744176.5299
RMSE	3121.5664
MAPE	19.3703

Dataset ini berasal dari Automobile (UCI Machine Learning Repository) yang memiliki 205 baris dan 26 fitur. Variabel price berkisar dari 5.118 hingga 45.400, artikel tentang dataset menyebut bahwa harga mobil dinyatakan dalam US dollar (USD). Nilai carlength (141,1–208,1) dan carwidth (60,3–72,3) merupakan ukuran bodi mobil dalam inci, enginesize merujuk pada kapasitas mesin dalam inchi kubik (atau cc), dan horsepower merupakan tenaga mesin dalam satuan horsepower.

Koefisien regresi yang diperoleh sekitar 195,33 untuk carlength, 90,94 untuk carwidth, 65,73 untuk enginesize dan 66,89 untuk horsepower. Artinya, setiap penambahan panjang mobil 1 inci menaikkan harga sekitar 195 USD. Lebar mobil juga berpengaruh positif (≈ 91 USD per inci). Kapasitas mesin dan tenaga mesin mempengaruhi harga dengan besaran serupa (~66–67 USD per satuan). Karakteristik bodi dan performa mesin jadi penentu utama harga, tetapi faktor lain seperti merek dan fitur interior tidak dimodelkan sehingga ketepatan prediksi masih terbatas.

3.3.5. Outlier



Dalam konteks regresi linier berganda, outlier merupakan titik data yang tidak mengikuti pola umum hubungan antara variabel bebas dan variabel terikat. Kehadiran outlier sering kali disebabkan oleh nilai ekstrim pada salah satu variabel atau kondisi khusus yang tidak tercakup oleh model, sehingga nilai residualnya jauh berbeda dibanding observasi lain.

Pada Gambar ditampilkan Aktual vs Prediksi dan grafik residual untuk Car Price Dataset. Garis biru menunjukkan harga mobil aktual per minggu, sedangkan tanda silang oranye menunjukkan prediksi model. Secara umum, prediksi mengikuti fluktuasi harga mobil, tetapi terdapat beberapa titik yang menyimpang jelas. Misalnya, lonjakan tajam pada minggu ke-17 sampai ke-19 harga mobil melonjak hingga sekitar 41 000 dolar,

membuat beberapa residual positif sangat besar. Pada minggu-minggu tersebut, model meremehkan harga mobil karena adanya unit yang memiliki spesifikasi luar biasa, seperti mesin berkapasitas besar atau fitur mewah yang tidak umum terdapat pada sampel lain. Sebaliknya, beberapa titik prediksi yang berada jauh di bawah nilai aktual menghasilkan residual negatif tajam, menandakan model terlalu tinggi memprediksi harga unit tertentu.

Grafik residual di bagian bawah membantu kita mengenali outlier tersebut. Titik-titik residual yang jauh berada di atas atau di bawah garis nol menandakan observasi yang dapat dikategorikan sebagai outlier. Dengan mengetahui posisi dan penyebab outlier, kita dapat mempertimbangkan langkah perbaikan seperti transformasi data, penghapusan observasi tertentu, atau penambahan variabel baru agar model lebih robust terhadap nilai ekstrim.

3.3.6. *Source Code*

```
%====== Multiple Regresi Linier ======
A = [50.0,  8512.3,  3272.5,  6370.0,  5264.0;
8512.3,  1460891.29,  558743.89,  1115424.5,
921589.0;
3272.5,  558743.89,  214504.71,  421508.5,
348495.6;
6370.0,  1115424.5,  421508.5,  942710.0,  771326.0;
5264.0,  921589.0,  348495.6,  771326.0,  642194.0]
;

B = [715226.167; 127932424.6594; 47723006.2393;
112941169.8770; 93128379.7200] ;

AB = [A, B];
[n, ~] = size(AB);
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
```

```
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);

X1 = [168.8; 168.8; 171.2; 176.6; 176.6; 177.3;
192.7; 192.7; 192.7; 178.2; 176.8; 176.8; 176.8;
176.8; 189.0; 189.0; 193.8; 197.0; 141.1; 155.9;
158.8; 157.3; 157.3; 157.3; 157.3; 157.3; 157.3;
157.3; 174.6; 173.2; 144.6; 144.6; 150.0; 150.0;
150.0; 163.4; 157.1; 167.5; 167.5; 175.4; 175.4;
175.4; 169.1; 170.7; 155.9; 155.9; 172.6; 199.6;
199.6; 191.7] ;
```

```
X2 = [64.1; 64.1; 65.5; 66.2; 66.4; 66.3; 71.4;
71.4; 71.4; 67.9; 64.8; 64.8; 64.8; 64.8; 66.9;
66.9; 67.9; 70.9; 60.3; 63.6; 63.6; 63.8; 63.8;
63.8; 63.8; 63.8; 63.8; 63.8; 64.6; 66.3; 63.9;
63.9; 64.0; 64.0; 64.0; 64.0; 63.9; 65.2; 65.2;
65.2; 62.5; 65.2; 66.0; 61.8; 63.6; 63.6; 65.2;
69.6; 69.6; 70.6] ;
X3 = [130.0; 130.0; 152.0; 109.0; 136.0; 136.0;
136.0; 136.0; 131.0; 131.0; 108.0; 108.0; 164.0;
164.0; 164.0; 209.0; 209.0; 209.0; 61.0; 90.0; 90.0;
90.0; 90.0; 98.0; 90.0; 90.0; 90.0; 98.0; 122.0;
156.0; 92.0; 92.0; 79.0; 92.0; 92.0; 92.0; 92.0;
110.0; 110.0; 110.0; 110.0; 110.0; 110.0; 111.0;
90.0; 90.0; 119.0; 258.0; 258.0; 326.0] ;
X4 = [111.0; 111.0; 154.0; 102.0; 115.0; 110.0;
110.0; 110.0; 140.0; 160.0; 101.0; 101.0; 121.0;
121.0; 121.0; 182.0; 182.0; 182.0; 48.0; 70.0; 70.0;
68.0; 68.0; 102.0; 68.0; 68.0; 68.0; 102.0; 88.0;
145.0; 58.0; 76.0; 60.0; 76.0; 76.0; 76.0; 76.0;
86.0; 86.0; 86.0; 86.0; 101.0; 100.0; 78.0; 70.0;
70.0; 90.0; 176.0; 176.0; 262.0] ;
Y  = [13495.0; 16500.0; 16500.0; 13950.0; 17450.0;
15250.0; 17710.0; 18920.0; 23875.0; 17859.167;
16430.0; 16925.0; 20970.0; 21105.0; 24565.0;
30760.0; 41315.0; 36880.0; 5151.0; 6295.0; 6575.0;
5572.0; 6377.0; 7957.0; 6229.0; 6692.0; 7609.0;
8558.0; 8921.0; 12964.0; 6479.0; 6855.0; 5399.0;
6529.0; 7129.0; 7295.0; 7295.0; 7895.0; 9095.0;
```

```
8845.0; 10295.0; 12945.0; 10345.0; 6785.0; 8916.5;
8916.5; 11048.0; 32250.0; 35550.0; 36000.0] ;

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];

Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);

% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y aktual','Y
prediksi','Location','northeast');
```

```
subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

3.3.7. Penjelasan *Source Code*

1. Menyusun sistem normal equations (A, B)

```
%====== Multiple Regresi Linier ======
A = [50.0,  8512.3,  3272.5,  6370.0,  5264.0;
8512.3,  1460891.29,  558743.89,  1115424.5,
921589.0;
3272.5,  558743.89,  214504.71,  421508.5,
348495.6;
6370.0,  1115424.5,  421508.5,  942710.0,  771326.0;
5264.0,  921589.0,  348495.6,  771326.0,  642194.0]
;

B = [715226.167; 127932424.6594; 47723006.2393;
112941169.8770; 93128379.7200] ;

AB = [A, B];
[n, ~] = size(AB);
```

Bagian ini memuat matriks normal equations hasil $X^T X$ (disimpan sebagai A) dan vektor $X^T y$ (disimpan sebagai B). Keduanya digabung menjadi matriks teraugmentasi AB = [A B] berukuran 5×6. Nilai n = 5 adalah banyaknya koefisien yang ingin dicari: b_0, b_1, b_2, b_3, b_4.

2. Eliminasi Gauss dengan partial pivoting

```
for i = 1:n
```

```
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
```

Partial pivoting: memilih elemen terbesar (secara absolut) pada kolom ke-i di baris i..n sebagai pivot, lalu menukar baris bila perlu. Tujuannya menstabilkan perhitungan dan mengurangi pembagian dengan nilai kecil.

Eliminasi maju: untuk setiap baris di bawah pivot (indeks j), dihitung faktor Factor = a_{ji}/a_{ii} kemudian dikurangi baris j dengan factor × baris i. Hasilnya, elemen di bawah pivot menjadi nol dan terbentuk matriks segitiga atas.

3. Substitusi balik → mendapatkan koefisien *b*

```
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
```

```
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);
```

Setelah bentuk segitiga atas diperoleh, dilakukan back-substitution: untuk baris i dihitung jumlah kontribusi kolom kanan $s = \sum_{kk=i+1}^{n} a_{ik} b_{k`}$ lalu $b_i = \frac{a_{i,n+1} - s}{a_{ii}}$. Hasil akhirnya adalah vektor koefisien regresi $b = [b_0, \dots, b_4]^T$.

4. Menyusun data asli dan matriks desain *X*

```
X1 = [168.8; 168.8; 171.2; 176.6; 176.6; 177.3;
192.7; 192.7; 192.7; 178.2; 176.8; 176.8; 176.8;
176.8; 189.0; 189.0; 193.8; 197.0; 141.1; 155.9;
158.8; 157.3; 157.3; 157.3; 157.3; 157.3; 157.3;
157.3; 174.6; 173.2; 144.6; 144.6; 150.0; 150.0;
150.0; 163.4; 157.1; 167.5; 167.5; 175.4; 175.4;
175.4; 169.1; 170.7; 155.9; 155.9; 172.6; 199.6;
199.6; 191.7] ;
X2 = [64.1; 64.1; 65.5; 66.2; 66.4; 66.3; 71.4;
71.4; 71.4; 67.9; 64.8; 64.8; 64.8; 64.8; 66.9;
66.9; 67.9; 70.9; 60.3; 63.6; 63.6; 63.8; 63.8;
63.8; 63.8; 63.8; 63.8; 63.8; 64.6; 66.3; 63.9;
63.9; 64.0; 64.0; 64.0; 64.0; 63.9; 65.2; 65.2;
65.2; 62.5; 65.2; 66.0; 61.8; 63.6; 63.6; 65.2;
69.6; 69.6; 70.6] ;
X3 = [130.0; 130.0; 152.0; 109.0; 136.0; 136.0;
136.0; 136.0; 131.0; 131.0; 108.0; 108.0; 164.0;
164.0; 164.0; 209.0; 209.0; 209.0; 61.0; 90.0; 90.0;
90.0; 90.0; 98.0; 90.0; 90.0; 90.0; 98.0; 122.0;
156.0; 92.0; 92.0; 79.0; 92.0; 92.0; 92.0; 92.0;
110.0; 110.0; 110.0; 110.0; 110.0; 110.0; 111.0;
90.0; 90.0; 119.0; 258.0; 258.0; 326.0] ;
```

```
X4 = [111.0; 111.0; 154.0; 102.0; 115.0; 110.0;
110.0; 110.0; 140.0; 160.0; 101.0; 101.0; 121.0;
121.0; 121.0; 182.0; 182.0; 182.0; 48.0; 70.0; 70.0;
68.0; 68.0; 102.0; 68.0; 68.0; 68.0; 102.0; 88.0;
145.0; 58.0; 76.0; 60.0; 76.0; 76.0; 76.0; 76.0;
86.0; 86.0; 86.0; 86.0; 101.0; 100.0; 78.0; 70.0;
70.0; 90.0; 176.0; 176.0; 262.0] ;
Y  = [13495.0; 16500.0; 16500.0; 13950.0; 17450.0;
15250.0; 17710.0; 18920.0; 23875.0; 17859.167;
16430.0; 16925.0; 20970.0; 21105.0; 24565.0;
30760.0; 41315.0; 36880.0; 5151.0; 6295.0; 6575.0;
5572.0; 6377.0; 7957.0; 6229.0; 6692.0; 7609.0;
8558.0; 8921.0; 12964.0; 6479.0; 6855.0; 5399.0;
6529.0; 7129.0; 7295.0; 7295.0; 7895.0; 9095.0;
8845.0; 10295.0; 12945.0; 10345.0; 6785.0; 8916.5;
8916.5; 11048.0; 32250.0; 35550.0; 36000.0] ;
m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];
```

- X1 menyimpan data Car Length, yaitu panjang mobil dalam satuan inci untuk setiap observasi.
- X2 berisi lebar mobil (Car Width) dalam inci.
- X3 merupakan kapasitas mesin (Engine Size) dalam satuan kubik atau cc. Nilai-nilainya menunjukkan besar kecilnya mesin yang dipasang.
- X4 memuat tenaga kuda (Horsepower) dari mesin setiap mobil. Keempat vektor tersebut masing-masing terdiri dari 50 baris karena dataset ini mempunyai 50 sampel mobil.

Y adalah vektor kolom yang memuat harga jual mobil dalam dolar (atau satuan yang setara). Nilai-nilai di dalamnya merupakan target yang ingin diprediksi oleh model regresi berdasarkan keempat fitur di atas.

m = rows(Y); digunakan untuk mendapatkan jumlah baris dari vektor Y, yang sama dengan jumlah sampel. Variabel m ini kemudian dipakai untuk membangun matriks desain dan menghitung metrik-metrik evaluasi.

X = [ones(m,1), X1, X2, X3, X4]; menghasilkan sebuah matriks berukuran m×5 yang terdiri dari kolom pertama berisi angka 1 (untuk mengakomodasi intercept b0) dan empat kolom berikutnya berisi masing-masing vektor X1 hingga X4. Susunan ini memastikan bahwa setiap baris matriks X mewakili satu mobil lengkap dengan empat fitur numeriknya, sehingga algoritma regresi dapat menghitung koefisien b0 – b4yang memodelkan hubungan linear antara spesifikasi mobil dan harga jual.

5. Prediksi dan metrik galat

```
Yprediksi = X * b;
e     = Y - Yprediksi;
SSE   = sum(e.^2);
MSE   = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);
```

$\hat{Y} = Xb$ adalah prediksi, Residual e = Y - $\hat{Y}$, MSE = SSE dibagi jumlah observasi, RMSE = akar dari MSE → satuannya sama dengan Y, MAPE menyatakan rata-rata persentase kesalahan absolut (dalam %). Penulisan %%%% pada sprintf/printf memang diperlukan untuk mencetak tanda % di output.

6. Grafik

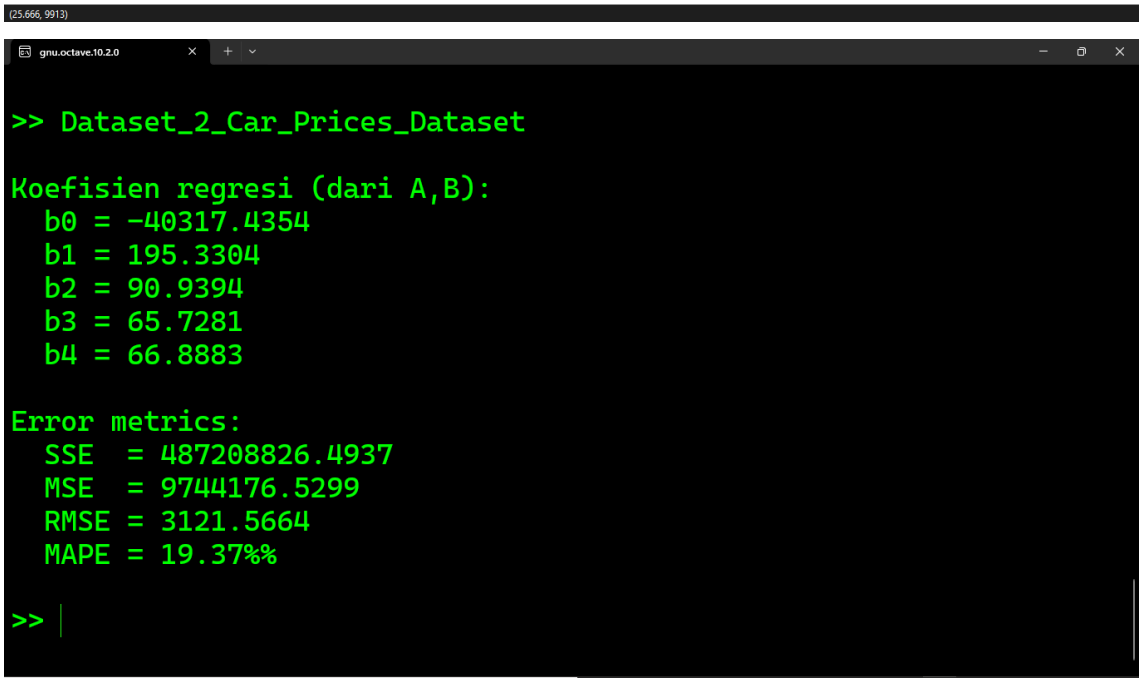
```
% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
```

```
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y                                    aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

- Subplot atas: membandingkan kurva aktual dan prediksi untuk melihat kedekatan pola.
- Subplot bawah: menampilkan residual per minggu. Titik yang jauh dari garis nol dapat diinterpretasi sebagai kandidat outlier atau observasi yang kurang terjelaskan oleh model.

  Di title(sprintf(...)) tutup format MAPE dengan %% agar tanda persen tercetak.

3.3.8. Tampilan Grafik dan Terminal



3.4. Implementasi Metode Multiple Regresi Linier Terhadap Medical Cost Personal Dataset

3.4.1. Tahap Pengerjaan

Membentuk matriks desain X dengan ukuran $m \times (n+1)$, di mana m adalah jumlah observasi dan n jumlah variabel bebas. Kolom pertama diisi dengan angka 1 untuk merepresentasikan intersep, sedangkan kolom berikutnya diisi variabel bebas (X1–X4).

Menyusun vektor target Y dengan panjang m yang memuat variabel terikat pada setiap observasi.

Sigma (X1)	Sigma (X2)	Sigma (X3)	Sigma (X4)	Sigma (Y)	Sigma X1^2
1844	1541.555	43	11.00	447425440	79492

Sigma (X2^2)	Sigma (X3^3)	Sigma (X4^2)
49246.23328	101	11

Sigma (X1X2)	Sigma (X1X3)	Sigma (X1X4)	Sigma (X2X3)	Sigma (X2X4)	Sigma (X3X4)
57341.875	1545	384	1344.72	383.61	6

Sigma (X1Y)	Sigma (X2Y)	Sigma (X3Y)	Sigma (X4Y)
16308407360	13734439590	380131500	97933500

3.4.2. Menyusun Matriks Augmented

$$\begin{bmatrix} N & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum (X1^2) & \sum (X2.X1) & \sum (X3.X1) & \sum (X4.X1) \\ \sum X2 & \sum (X1.X2) & \sum (X2^2) & \sum (X3.X2) & \sum (X4.X2) \\ \sum X3 & \sum (X1X3) & \sum (X2.X3) & \sum (X3^2) & \sum (X4.X3) \\ \sum X4 & \sum (X1.X4) & \sum (X2.X4) & \sum (X3.X4) & \sum (X4^2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \sum (Y) \\ \sum (X1.Y) \\ \sum (X2.Y) \\ \sum (X3.Y) \\ \sum (X4.\ Y) \end{bmatrix}$$

Membentuk normal equation dengan menghitung $A = X^T X$ dan $B = X^T Y$. Matriks A berukuran $(n+1) \times (n+1)$ dan vektor B berukuran $(n+1) \times 1$.

Matriks Augmented					
50	1844	1541.555	43	11.00	447425440
1844	79492	57341.875	1545	384	16308407360

1541.555	57341.875	49246.23328	1344.72	383.61	13734439590
43	1545	1344.72	101	6	380131500
11.00	384	383.61	6	11	97933500

3.4.3. Tranformasi ke Bentuk Segitiga Atas

Eliminasi Gauss: Matriks augmented $[AB]$ direduksi menjadi bentuk segitiga atas menggunakan eliminasi baris elementer. Proses pivoting dilakukan untuk menghindari pembagian oleh angka mendekati nol. Penulis membagi seluruh baris pertama dengan 50 supaya elemen pertama menjadi 1.

$$R_1 \leftarrow \frac{1}{50} R_1$$

Matriks Bentuk Segitiga Atas Tahap 1 (R1 <- 1/50R1)					
1	36.88	30.8311	0.86	0.22	8948508.8
1844	79492	57341.875	1545	384	16308407360
1541.555	57341.875	49246.23328	1344.72	383.61	13734439590
43	1545	1344.72	101	6	380131500
11.00	384	383.61	6	11	97933500

Penulis menginginkan kolom pertama menjadi seperti:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Jadi penulis melakukan:

$$R_2 \leftarrow R_2 - 1844R_1$$
$$R_3 \leftarrow R_3 - 1541.555R_1$$
$$R_4 \leftarrow R_4 - 43R_1$$
$$R_5 \leftarrow R_5 - 11R_1$$

Setelah eliminasi bentuknya menjadi:

Matriks Bentuk Segitiga Atas Tahap (R2 <-R2 - 1844R1, R3 <-R3 - 1541.555R1, R4 <-R4 - 43R1, R5 <-R5 - 11R1)					
1	36.88	30.8311	0.86	0.22	8948508.8
0	11485.28	489.3266	-40.84	-21.68	-192642867.2
0	489.3266	1718.396915	18.9827	44.4679	-60178893.68
0	-40.84	18.9827	64.02	-3.46	-4654378.4
0	-21.68	44.4679	-3.46	8.58	-500096.8

Matriks Bentuk Segitiga Atas Tahap ($R_3 \leftarrow R_3 - 489.3266R_2$, $R_4 \leftarrow R_4 - (-40.84)R_2$, $R_5 \leftarrow R_5 - (-21.68)R_2$)					
1	36.88	30.8311	0.86	0.22	8948508.8
0	1	0.0426	-0.0036	-0.0019	-16773.0231
0	0	1697.5493	20.7227	45.3916	-51971407.2996
0	0	20.7227	63.8748	-3.5371	-5339388.6650
0	0	45.3916	-3.5371	8.5391	-863735.9417

Matriks Bentuk Segitiga Atas Tahap (Normalisasi)					
1	36.88	30.8311	0.86	0.22	8948508.8
0	1	0.0426	-0.0036	-0.0019	-16773.0231
0	0	1	0.0122	0.0267	-30615.5508
0	0	0	63.6218	-4.0912	-4704952.5602
0	0	0	-4.0912	7.3253	525951.9551

Matriks Bentuk Segitiga Atas Tahap (Normalisasi R4)					
1	36.88	30.8311	0.86	0.22	8948508.8
0	1	0.0426	-0.0036	-0.0019	-16773.0231

0	0	0	0.0000	0.0000	-18.0351
0	0	0	1	-0.0643	-73951.8831
0	0	0	0	7.0622	223399.6930

Setelah matriks berada pada bentuk segitiga atas, koefisien regresi diperoleh b_0, b_1, b_2, b_3, b_4 dengan menyelesaikan sistem persamaan linear dari baris paling bawah ke atas.

Substitusi Mundur	
b0	10524084.4572
b1	-15666.0420
b2	-30583.4705
b3	-71917.7237
b4	31632.9596

	Y Prediksi	e = Y - Y Prediksi
1	9404783.7921	3895216.2079
2	9137374.1791	3112625.8209
3	8860427.5840	3389572.4160
4	9312707.3738	2902292.6262
5	9139520.4855	2270479.5145
6	9251218.6248	1598781.3752
7	8708817.5484	1441182.4516
8	8880302.2608	1269697.7392
9	8888300.5312	981699.4688

10	8793845.0598	1006154.9402
11	9330534.8110	469465.1890
12	8780383.3737	900616.6263
13	9111694.1064	198305.8936
14	8428952.3104	811047.6896
15	8844252.6710	395747.3290
16	9402158.5615	-302158.5615
17	8696173.3279	403826.6721
18	9434502.6374	-474502.6374
19	8414272.2445	475727.7555
20	9006139.6485	-151139.6485
21	8482964.0823	267035.9177
22	8991281.0297	-311281.0297
23	9199199.3575	-554199.3575
24	8974929.8871	-329929.8871
25	8943503.6954	-368503.6954
26	8536261.0063	3738.9937
27	8831104.3950	-368104.3950
28	8516243.4546	-116243.4546
29	9560154.1331	-1160154.1331
30	8816054.6886	-416054.6886
31	9122292.9433	-722292.9433
32	9437291.6753	-1037291.6753
33	8992153.7848	-697153.7848
34	8671305.7617	-481305.7617
35	8931912.1913	-811912.1913

36	9601762.2745	-1520822.2745
37	8328852.5776	-285852.5776
38	9480631.1790	-1500631.1790
39	8813992.3602	-851492.3602
40	8395474.4243	-485474.4243
41	9334579.1342	-1459579.1342
42	8774329.1838	-934329.1838
43	9143751.0243	-1443751.0243
44	8858634.5648	-1158634.5648
45	8723739.5559	-1163739.5559
46	8521688.6980	-961688.6980
47	8915750.3673	-1390750.3673
48	9022048.0123	-1532048.0123
49	8833909.4062	-1378909.4062
50	8843284.0199	-1423284.0199

3.4.4. Nilai Error

Vektor koefisien b digunakan untuk menghitung prediksi dengan rumus $\hat{Y} = Xb$,. Selisih antara nilai aktual dan prediksi Y - $\hat{Y}$ membentuk residual yang dianalisis untuk mengidentifikasi outlier.

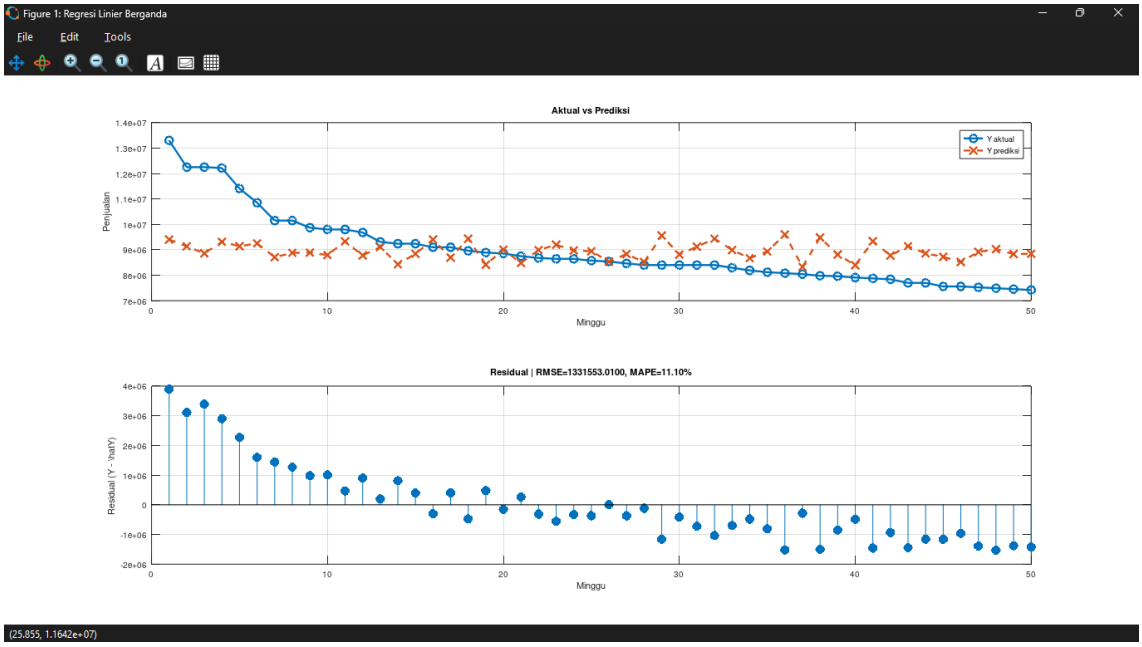
RSS	88651670924548.0000
MSE	1773033418490.9600
RMSE	1331553.0100
MAPE	11.1000

Proyek GitHub mengenai dataset ini menjelaskan bahwa charges merupakan "medical costs billed by health insurance (USD)". Dengan demikian, premi asuransi

dinyatakan dalam dolar AS. Variabel age diukur dalam tahun, BMI dalam kg/m², children sebagai jumlah anggota tanggungan, dan smoker dikodekan 1 untuk perokok dan 0 untuk bukan perokok.

Koefisien regresi menunjukkan perokok membayar premi jauh lebih tinggi (~31.633 USD) dibanding non-perokok. Hasil sampel 50 baris yang dianalisis memunculkan koefisien negatif untuk age dan BMI, yang bertentangan dengan ekspektasi (premi biasanya meningkat dengan usia dan BMI). Hal ini kemungkinan disebabkan oleh kolinearitas atau variabilitas yang terbatas pada sampel. Secara umum, hubungan positif antara perokok dan premi konsisten dengan literatur, status perokok menjadi prediktor paling kuat terhadap besarnya biaya asuransi.

3.4.5. Outlier



3.4.6. *Source Code*

```
%====== Multiple Regresi Linier ======
A = [50.0,  1844.0,  1541.555,  43.0,  11.0;
1844.0,  79492.0,  57341.875,  1545.0,  384.0;
1541.555,  57341.875,  49246.233275,  1344.72,
383.61;
```

```
43.0,  1545.0,  1344.72,  101.0,  6.0;
11.0,  384.0,  383.61,  6.0,  11.0] ;

B = [447425440.0; 16308407360.0; 13734439589.5;
380131500.0; 97933500.0] ;

AB = [A, B];
[n, ~] = size(AB);
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
```

```
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);

X1 = [19.0; 18.0; 28.0; 33.0; 32.0; 31.0; 46.0;
37.0; 37.0; 60.0; 25.0; 62.0; 23.0; 56.0; 27.0;
19.0; 52.0; 23.0; 56.0; 30.0; 60.0; 30.0; 18.0;
34.0; 37.0; 59.0; 63.0; 55.0; 23.0; 31.0; 22.0;
18.0; 19.0; 63.0; 28.0; 19.0; 62.0; 26.0; 35.0;
60.0; 24.0; 31.0; 41.0; 37.0; 38.0; 55.0; 18.0;
28.0; 60.0; 36.0] ;
X2 = [27.9; 33.77; 33.0; 22.705; 28.88; 25.74;
33.44; 27.74; 29.83; 25.84; 26.22; 26.29; 34.4;
39.82; 42.13; 24.6; 30.78; 23.845; 40.3; 35.3;
36.005; 32.4; 34.1; 31.92; 28.025; 27.72; 23.085;
32.775; 17.385; 36.3; 35.6; 26.315; 28.6; 28.31;
36.4; 20.425; 32.965; 20.8; 36.67; 39.9; 26.6;
36.63; 21.78; 30.8; 37.05; 37.3; 38.665; 34.77;
24.53; 35.2] ;
X3 = [0.0; 1.0; 3.0; 0.0; 0.0; 0.0; 1.0; 3.0; 2.0;
0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 1.0; 1.0; 0.0; 0.0;
0.0; 0.0; 1.0; 0.0; 1.0; 2.0; 3.0; 0.0; 2.0; 1.0;
2.0; 0.0; 0.0; 5.0; 0.0; 1.0; 0.0; 3.0; 0.0; 1.0;
0.0; 0.0; 2.0; 1.0; 2.0; 1.0; 0.0; 2.0; 0.0; 0.0;
1.0] ;
X4 = [1.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
0.0; 0.0; 1.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 0.0;
1.0; 0.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 0.0; 0.0;
1.0; 1.0; 0.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 1.0;
```

```
1.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
1.0] ;
Y  = [13300000.0; 12250000.0; 12250000.0;
12215000.0; 11410000.0; 10850000.0; 10150000.0;
10150000.0; 9870000.0; 9800000.0; 9800000.0;
9681000.0; 9310000.0; 9240000.0; 9240000.0;
9100000.0; 9100000.0; 8960000.0; 8890000.0;
8855000.0; 8750000.0; 8680000.0; 8645000.0;
8645000.0; 8575000.0; 8540000.0; 8463000.0;
8400000.0; 8400000.0; 8400000.0; 8400000.0;
8400000.0; 8295000.0; 8190000.0; 8120000.0;
8080940.0; 8043000.0; 7980000.0; 7962500.0;
7910000.0; 7875000.0; 7840000.0; 7700000.0;
7700000.0; 7560000.0; 7560000.0; 7525000.0;
7490000.0; 7455000.0; 7420000.0];

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];

Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
```

```
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);

% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

3.4.7. Penjelasan *Source Code*

1. Menyusun sistem normal equations (A, B)

```
%====== Multiple Regresi Linier ======
A = [50.0,  1844.0,  1541.555,  43.0,  11.0;
1844.0,  79492.0,  57341.875,  1545.0,  384.0;
1541.555,  57341.875,  49246.233275,  1344.72,
383.61;
43.0,  1545.0,  1344.72,  101.0,  6.0;
11.0,  384.0,  383.61,  6.0,  11.0] ;
```

```
B = [447425440.0; 16308407360.0; 13734439589.5;
380131500.0; 97933500.0] ;

AB = [A, B];
[n, ~] = size(AB);
```

Bagian ini memuat matriks normal equations hasil $X^T X$ (disimpan sebagai A) dan vektor $X^T y$ (disimpan sebagai B). Keduanya digabung menjadi matriks teraugmentasi AB = [A B] berukuran 5×6. Nilai n = 5 adalah banyaknya koefisien yang ingin dicari: b_0, b_1, b_2, b_3, b_4.

2. Eliminasi Gauss dengan partial pivoting

```
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
```

Partial pivoting: memilih elemen terbesar (secara absolut) pada kolom ke-i di baris i..n sebagai pivot, lalu menukar baris bila perlu. Tujuannya menstabilkan perhitungan dan mengurangi pembagian dengan nilai kecil.

Eliminasi maju: untuk setiap baris di bawah pivot (indeks j), dihitung faktor Factor = a_{ji}/a_{ii} kemudian dikurangi baris j dengan factor × baris i. Hasilnya, elemen di bawah pivot menjadi nol dan terbentuk matriks segitiga atas.

3. Substitusi balik → mendapatkan koefisien *b*

```
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);
```

Setelah bentuk segitiga atas diperoleh, dilakukan back-substitution: untuk baris i dihitung jumlah kontribusi kolom kanan $s = \sum_{kk=i+1}^{n} a_{ik} b_{k`}$ lalu $b_i = \frac{a_{i,n+1} - s}{a_{ii}}$. Hasil akhirnya adalah vektor koefisien regresi $b = [b_0, \dots, b_4]^T$.

4. Menyusun data asli dan matriks desain *X*

```
X1 = [19.0; 18.0; 28.0; 33.0; 32.0; 31.0; 46.0;
37.0; 37.0; 60.0; 25.0; 62.0; 23.0; 56.0; 27.0;
19.0; 52.0; 23.0; 56.0; 30.0; 60.0; 30.0; 18.0;
34.0; 37.0; 59.0; 63.0; 55.0; 23.0; 31.0; 22.0;
18.0; 19.0; 63.0; 28.0; 19.0; 62.0; 26.0; 35.0;
60.0; 24.0; 31.0; 41.0; 37.0; 38.0; 55.0; 18.0;
28.0; 60.0; 36.0] ;
X2 = [27.9; 33.77; 33.0; 22.705; 28.88; 25.74;
33.44; 27.74; 29.83; 25.84; 26.22; 26.29; 34.4;
39.82; 42.13; 24.6; 30.78; 23.845; 40.3; 35.3;
36.005; 32.4; 34.1; 31.92; 28.025; 27.72; 23.085;
32.775; 17.385; 36.3; 35.6; 26.315; 28.6; 28.31;
36.4; 20.425; 32.965; 20.8; 36.67; 39.9; 26.6;
```

```
36.63; 21.78; 30.8; 37.05; 37.3; 38.665; 34.77;
24.53; 35.2] ;
X3 = [0.0; 1.0; 3.0; 0.0; 0.0; 0.0; 1.0; 3.0; 2.0;
0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 1.0; 1.0; 0.0; 0.0;
0.0; 0.0; 1.0; 0.0; 1.0; 2.0; 3.0; 0.0; 2.0; 1.0;
2.0; 0.0; 0.0; 5.0; 0.0; 1.0; 0.0; 3.0; 0.0; 1.0;
0.0; 0.0; 2.0; 1.0; 2.0; 1.0; 0.0; 2.0; 0.0; 0.0;
1.0] ;
X4 = [1.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
0.0; 0.0; 1.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 0.0;
1.0; 0.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 0.0; 0.0;
1.0; 1.0; 0.0; 0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 1.0;
1.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0;
1.0] ;
Y  = [13300000.0; 12250000.0; 12250000.0;
12215000.0; 11410000.0; 10850000.0; 10150000.0;
10150000.0; 9870000.0; 9800000.0; 9800000.0;
9681000.0; 9310000.0; 9240000.0; 9240000.0;
9100000.0; 9100000.0; 8960000.0; 8890000.0;
8855000.0; 8750000.0; 8680000.0; 8645000.0;
8645000.0; 8575000.0; 8540000.0; 8463000.0;
8400000.0; 8400000.0; 8400000.0; 8400000.0;
8400000.0; 8295000.0; 8190000.0; 8120000.0;
8080940.0; 8043000.0; 7980000.0; 7962500.0;
7910000.0; 7875000.0; 7840000.0; 7700000.0;
7700000.0; 7560000.0; 7560000.0; 7525000.0;
7490000.0; 7455000.0; 7420000.0];

m  = rows(Y);
```

```
X  = [ones(m,1), X1, X2, X3, X4];
```

- X1 berisi data Age (usia) dari masing-masing peserta asuransi. Usia dicatat dalam tahun dengan 50 observasi.
- X2 memuat nilai BMI (Body Mass Index) peserta, yang mencerminkan tingkat berat badan relatif terhadap tinggi badan.
- X3 menunjukkan jumlah anak tanggungan (Children) yang dimiliki peserta. Nilai 0–5 menunjukkan berapa banyak anak dalam asuransi mereka.
- X4 adalah variabel kategori yang telah dikodekan menjadi numerik (1 untuk perokok dan 0 untuk non-perokok). Karena status merokok sangat mempengaruhi premi, variabel ini penting dalam model.

Vektor Y menyimpan biaya asuransi (Charges) yang dibayar oleh setiap peserta, dinyatakan dalam rupiah. Nilai-nilainya berkisar antara sekitar 7,4 juta hingga 13,3 juta. Baris m = rows(Y); menghitung jumlah baris dari vektor Y, yang sama dengan banyaknya observasi (50 peserta). Variabel m ini diperlukan untuk membuat vektor konstanta dan memastikan ukuran matriks konsisten. Baris X = [ones(m,1), X1, X2, X3, X4]; membuat matriks desain berukuran m×5. Kolom pertama diisi dengan angka 1 sebagai konstanta untuk menghitung intercept $b0$. Kolom berikutnya adalah variabel usia, BMI, jumlah anak, dan status perokok. Setiap baris matriks X mewakili satu peserta asuransi dengan keempat variabel independennya. Matriks ini kemudian digunakan dalam perhitungan regresi linier berganda untuk memperoleh koefisien $b0$ – $b4$, yang merepresentasikan pengaruh masing-masing variabel terhadap biaya asuransi.

5. Prediksi dan metrik galat

```
Yprediksi = X * b;
e     = Y - Yprediksi;
SSE   = sum(e.^2);
MSE   = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;
```

```
printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);
```

$\hat{Y} = Xb$ adalah prediksi, Residual e = Y - $\hat{Y}$, MSE = SSE dibagi jumlah observasi, RMSE = akar dari MSE → satuannya sama dengan Y, MAPE menyatakan rata-rata persentase kesalahan absolut (dalam %). Penulisan %%%% pada sprintf/printf memang diperlukan untuk mencetak tanda % di output.

6. Grafik

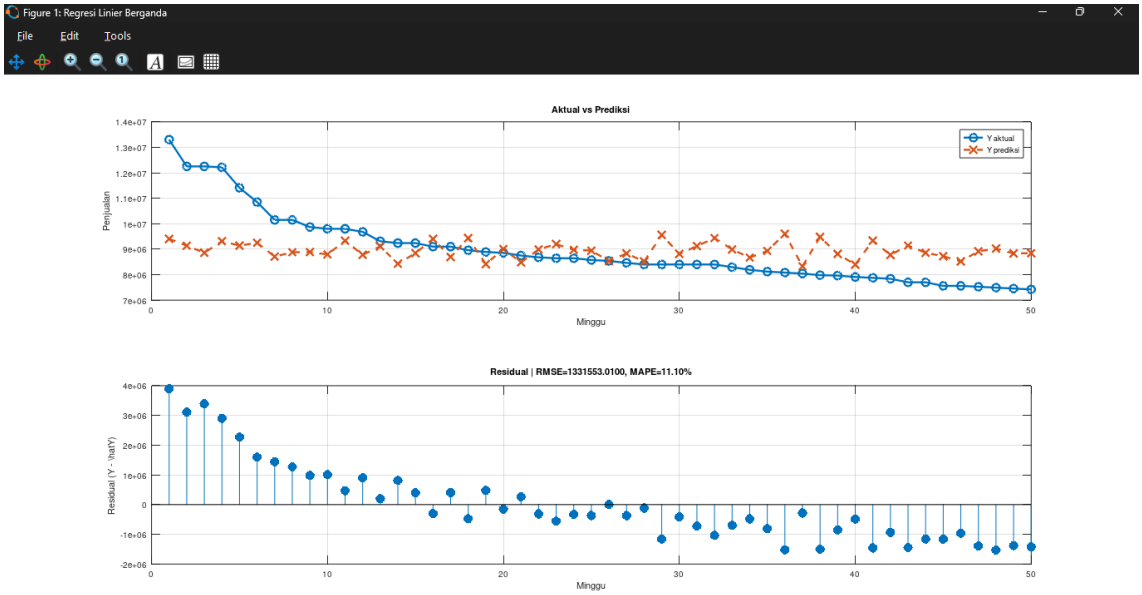
```
% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y                                    aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual  |  RMSE=%.4f,  MAPE=%.2f%%',
RMSE, MAPE));
```

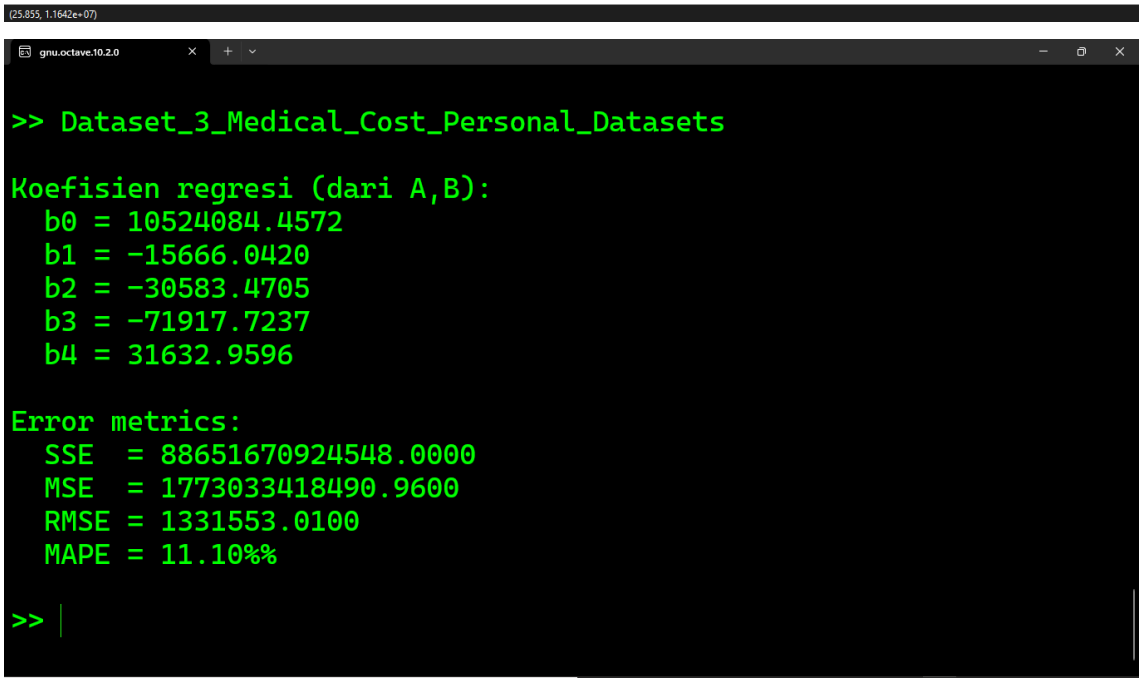
- Subplot atas: membandingkan kurva aktual dan prediksi untuk melihat kedekatan pola.

- Subplot bawah: menampilkan residual per minggu. Titik yang jauh dari garis nol dapat diinterpretasi sebagai kandidat outlier atau observasi yang kurang terjelaskan oleh model.

  Di title(sprintf(...)) tutup format MAPE dengan %% agar tanda persen tercetak.

3.4.8. Tampilan Grafik dan Terminal





3.5. Implementasi Metode Multiple Regresi Linier Terhadap Fish Market Dataset

3.5.1. Tahap Pengerjaan

Membentuk matriks desain X dengan ukuran m×(n+1), di mana m adalah jumlah observasi dan n jumlah variabel bebas. Kolom pertama diisi dengan angka 1 untuk merepresentasikan intersep, sedangkan kolom berikutnya diisi variabel bebas (X1–X4). Menyusun vektor target Y dengan panjang m yang memuat variabel terikat pada setiap observasi.

Sigma (X1)	Sigma (X2)	Sigma (X3)	Sigma (X4)	Sigma (Y)	Sigma X1^2
1352.4	1473.8	624.3352	240.74	23494	38342.52

Sigma (X2^2)	Sigma (X3^3)	Sigma (X4^2)
45593.56	8785.20356	1223.394672

Sigma (X1X2)	Sigma (X1X3)	Sigma (X1X4)	Sigma (X2X3)	Sigma (X2X4)	Sigma (X3X4)
41809.28	18164.31133	6837.80626	19822.34069	7458.3868	3248.063845

Sigma (X1Y)	Sigma (X2Y)	Sigma (X3Y)	Sigma (X4Y)
716978	782917.4	353813.0909	128681.0999

3.5.2. Menyusun Matriks Augmented

$$\begin{bmatrix} N & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum (X1^2) & \sum (X2.X1) & \sum (X3.X1) & \sum (X4.X1) \\ \sum X2 & \sum (X1.X2) & \sum (X2^2) & \sum (X3.X2) & \sum (X4.X2) \\ \sum X3 & \sum (X1X3) & \sum (X2.X3) & \sum (X3^2) & \sum (X4.X3) \\ \sum X4 & \sum (X1.X4) & \sum (X2.X4) & \sum (X3.X4) & \sum (X4^2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \sum (Y) \\ \sum (X1.Y) \\ \sum (X2.Y) \\ \sum (X3.Y) \\ \sum (X4.\ Y) \end{bmatrix}$$

Membentuk normal equation dengan menghitung $A = X^T X$ dan $B = X^T Y$. Matriks A berukuran $(n+1) \times (n+1)$ dan vektor B berukuran $(n+1) \times 1$.

Matriks Augmented					
50	1352.4	1473.8	624.3352	240.74	23494
1352.4	38342.52	41809.28	18164.31133	6837.80626	716978
1473.8	41809.28	45593.56	19822.34069	7458.3868	782917.4
624.3352	18164.31133	19822.34069	8785.203562	3248.063845	353813
240.74	6837.80626	7458.3868	3248.063845	1223.394672	128681.0999

3.5.3. Tranformasi ke Bentuk Segitiga Atas

Eliminasi Gauss: Matriks augmented $[AB]$ direduksi menjadi bentuk segitiga atas menggunakan eliminasi baris elementer. Proses pivoting dilakukan untuk menghindari pembagian oleh angka mendekati nol. Penulis membagi seluruh baris pertama dengan 50 supaya elemen pertama menjadi 1.

$$R_1 \leftarrow \frac{1}{50} R_1$$

Matriks Bentuk Segitiga Atas Tahap 1 (R1 <- 1/50R1)					
1	27.048	29.476	12.486704	4.81489	469.88
1352.4	38342.52	41809.28	18164.31133	6837.80626	716978
1473.8	41809.28	45593.56	19822.34069	7458.3868	782917.4
624.3352	18164.31133	19822.34069	8785.203562	3248.063845	353813.0909
240.74	6837.80626	7458.3868	3248.063845	1223.394672	128681.0999

Penulis menginginkan kolom pertama menjadi seperti:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Jadi penulis melakukan:

$$R_2 \leftarrow R_2 - 1352.4R_1$$

$$R_3 \leftarrow R_3 - 1473.8R_1$$

$$R_4 \leftarrow R_4 - 624.3352R_1$$

$$R_5 \leftarrow R_5 - 240.74R_1$$

Setelah eliminasi bentuknya menjadi:

Matriks Bentuk Segitiga Atas Tahap (R2 <-R2 - 1352.4R1, R3 <-R3 - 1473.8R1, R4 <-R4 - 624.3352R1, R5 <-R5 - 240.74R1)					
1	27.048	29.476	12.486704	4.81489	469.88
0	1762.8048	1945.9376	1277.29284	326.149024	81512.288
0	1945.9376	2151.8312	1419.436335	362.201918	90408.256
0	1277.29284	1419.436335	989.3147225	241.9585336	60450.46712
0	326.149024	362.201918	241.9585336	64.23638631	15560.07424

Matriks Bentuk Segitiga Atas Tahap ($R_3 \leftarrow R_3 - 1945.9376R_2$, $R_4 \leftarrow R_4 - 1277.29284R_2$, $R_5 \leftarrow R_5 - 326.149024R_2$)					
1	27.048	29.476	12.486704	4.81489	469.88
0	1	1.1039	0.7246	0.1850	46.2401
0	0	3.7357	9.4492	2.1702	427.8872
0	0	9.4492	63.8141	5.6375	1388.3056
0	0	2.1702	5.6375	3.8932	478.9075

Matriks Bentuk Segitiga Atas Tahap (Normalisasi)					
1	27.048	29.476	12.486704	4.81489	469.88
0	1	1.1039	0.7246	0.1850	46.2401
0	0	1	2.5295	0.5809	114.5415
0	0	0	39.9129	0.1481	305.9849
0	0	0	0.1481	2.6325	230.3300

Matriks Bentuk Segitiga Atas Tahap (Normalisasi R4)					
1	27.048	29.476	12.486704	4.81489	469.88
0	1	1.1039	0.7246	0.1850	46.2401
0	0	0	0.6771	0.1555	30.6617
0	0	0	1	0.0037	7.6663
0	0	0	0	2.6319	229.1946

Setelah matriks berada pada bentuk segitiga atas, koefisien regresi diperoleh b_0, b_1, b_2, b_3, b_4 dengan menyelesaikan sistem persamaan linear dari baris paling bawah ke atas.

Substitusi Mundur	
b0	-694.7704
b1	-25.2841
b2	45.3777
b3	7.3432
b4	87.0821

	Y Prediksi	e = Y - Y Prediksi
1	305.8945	-63.8945
2	358.4272	-68.4272
3	403.2863	-63.2863
4	437.6829	-74.6829
5	489.6111	-59.6111
6	504.3054	-54.3054

7	539.1177	-39.1177
8	470.1714	-80.1714
9	493.3670	-43.3670
10	514.0703	-14.0703
11	543.0874	-68.0874
12	511.0806	-11.0806
13	480.2694	19.7306
14	555.3487	-215.3487
15	574.0591	25.9409
16	613.2443	-13.2443
17	603.4420	96.5580
18	596.3573	103.6427
19	605.9605	4.0395
20	646.6292	3.3708
21	652.7727	-77.7727
22	639.2588	45.7412
23	648.0978	-28.0978
24	736.8675	-56.8675
25	692.8436	7.1564
26	736.6731	-11.6731
27	734.8341	-14.8341
28	742.8835	-28.8835
29	773.2999	76.7001
30	851.3934	148.6066
31	848.9400	71.0600
32	848.0406	106.9594

33	907.8562	17.1438
34	944.5004	30.4996
35	889.0869	60.9131
36	-153.1544	193.1544
37	-1.4587	70.4587
38	9.7218	68.2782
39	61.2312	25.7688
40	99.3926	20.6074
41	177.8095	-16.8095
42	107.1539	2.8461
43	99.4241	20.5759
44	157.3070	-7.3070
45	142.8779	2.1221
46	192.2302	-32.2302
47	132.9067	7.0933
48	170.6408	-10.6408
49	227.3470	-58.3470
50	177.8095	-16.8095

3.5.4. Nilai Error

Vektor koefisien b digunakan untuk menghitung prediksi dengan rumus $\hat{Y} = Xb$,. Selisih antara nilai aktual dan prediksi Y - $\hat{Y}$ membentuk residual yang dianalisis untuk mengidentifikasi outlier.

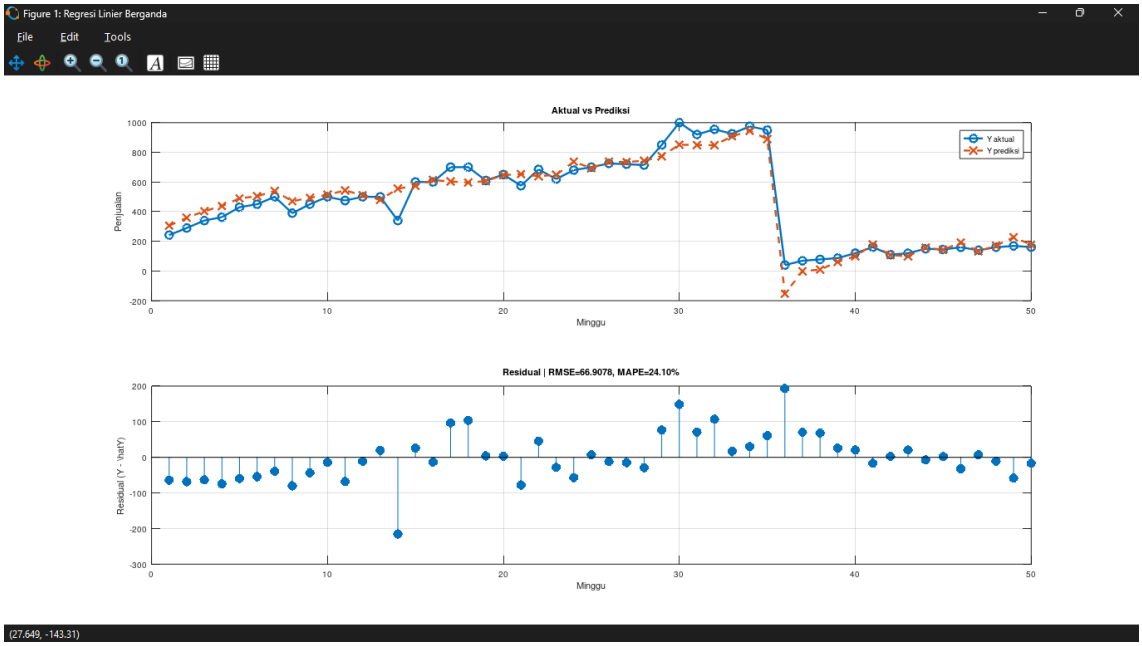
RSS	223832.7383
MSE	4476.6548
RMSE	66.9078

MAPE	24.0959

Dataset mencakup tujuh kolom, menurut kartu dataset, Weight diukur dalam gram. Sementara Length1 (vertical length), Length2 (diagonal length), Length3, Height, dan Width diukur dalam sentimeter.

Koefisien regresi menunjukkan bahwa lebar (Width) merupakan prediktor utama adalah kenaikan lebar 1 cm menambah berat sekitar 87 g. Panjang diagonal (Length2) menambah sekitar 45 g per cm, sementara tinggi (Height) memberi tambahan sekitar 7 g per cm. Koefisien untuk Length1 negatif (−25,28), menandakan adanya multikolineritas antara ukuran Panjang peningkatan dan panjang vertical dengan dimensi lain tetap konstan justru menurunkan prediksi berat. Singkatnya, dimensi lebar dan diagonal paling berpengaruh terhadap berat ikan, sementara hubungan panjang vertical kurang linier.

3.5.5. Outlier



Pada Fish Market Dataset, outlier merepresentasikan spesimen ikan yang memiliki berat jauh di atas atau di bawah berat rata-rata untuk ukuran tubuh tertentu. Karena model regresi memprediksi berat berdasarkan pengukuran fisik (panjang, tinggi, lebar), ikan yang sangat kurus atau sangat gemuk relatif terhadap dimensinya akan menghasilkan residual yang besar.

Grafik Aktual vs Prediksi pada Gambar menunjukkan tren kenaikan berat seiring bertambahnya ukuran ikan. Namun sekitar observasi ke-35 terdapat penurunan tajam dalam bobot aktual (berat ikan turun dari sekitar 900 gram menjadi 40 gram), sementara prediksi model masih tinggi karena berpedoman pada panjang ikan yang relatif besar. Hal ini menghasilkan residual negatif yang sangat besar pada minggu tersebut, seperti terlihat pada grafik residual. Setelah kejadian ini, kurva aktual kembali naik, tetapi beberapa titik masih menunjukkan deviasi signifikan. Titik-titik residual yang melampaui ±100 gram menandakan outlier, yaitu ikan dengan kombinasi panjang, lebar, dan tinggi yang tidak biasa sehingga tidak cocok dengan pola umum data. Analisis outlier ini penting agar mahasiswa dapat mempertimbangkan apakah model linier masih layak atau perlu transformasi/non-linier untuk menangkap hubungan antara dimensi tubuh dan berat ikan yang kompleks.

3.5.6. *Source Code*

```
%====== Multiple Regresi Linier ======
A = [50.0,  1352.4,  1473.8,  624.3352,  240.7445;
1352.4,  38342.52,  41809.28,  18164.31133,
6837.80626;
1473.8,  41809.28,  45593.56,  19822.34069,
7458.3868;
624.3352,  18164.31133,  19822.34069,  8785.203562,
3248.063845;
240.7445,  6837.80626,  7458.3868,  3248.063845,
1223.394672];

B = [23494.0; 716978.0; 782917.4; 353813.0909;
128681.0999];

AB = [A, B];
[n, ~] = size(AB);
```

```
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);

X1 = [23.2; 24.0; 23.9; 26.3; 26.5; 26.8; 26.8;
27.6; 27.6; 28.5; 28.4; 28.7; 29.1; 29.5; 29.4;
29.4; 30.4; 30.4; 30.9; 31.0; 31.3; 31.4; 31.5;
31.8; 31.9; 31.8; 32.0; 32.7; 32.8; 33.5; 35.0;
35.0; 36.2; 37.4; 38.0; 12.9; 16.5; 17.5; 18.2;
```

```
18.6; 22.0; 19.1; 19.4; 20.4; 20.5; 20.5; 21.0;
21.1; 22.0; 22.0];
X2 = [25.4; 26.3; 26.5; 29.0; 29.0; 29.7; 29.7;
30.0; 30.0; 30.7; 31.0; 31.0; 31.5; 32.0; 32.0;
32.0; 33.0; 33.0; 33.5; 33.5; 34.0; 34.0; 34.5;
35.0; 35.0; 35.0; 35.0; 36.0; 36.0; 37.0; 38.5;
38.5; 39.5; 41.0; 41.0; 14.1; 18.2; 18.8; 19.8;
20.0; 23.4; 20.8; 21.0; 22.0; 22.0; 22.5; 22.5;
22.5; 24.0; 23.4];
X3 = [11.52; 12.48; 12.3778; 12.73; 12.444; 13.6024;
14.1795; 12.67; 14.0049; 14.2266; 14.2628; 14.3714;
13.7592; 13.9129; 14.9544; 15.438; 14.8604; 14.938;
15.633; 14.4738; 15.1285; 15.9936; 15.5227; 15.4686;
16.2405; 16.36; 16.3618; 16.517; 16.8896; 18.957;
18.0369; 18.084; 18.7542; 18.6354; 17.6235; 4.1472;
5.2983; 5.5756; 5.6166; 6.216; 6.9153; 6.1677;
6.1146; 5.8045; 6.6339; 7.0334; 6.55; 6.4; 7.5344;
6.9153];
X4 = [4.02; 4.3056; 4.6961; 4.4555; 5.134; 4.9274;
5.2785; 4.69; 4.8438; 4.9594; 5.1042; 4.8146; 4.368;
5.0728; 5.1708; 5.58; 5.2854; 5.1975; 5.1338;
5.7276; 5.5695; 5.3704; 5.2801; 6.1306; 5.589;
6.0532; 6.09; 5.8515; 6.1984; 6.603; 6.3063; 6.292;
6.7497; 6.7473; 6.3705; 2.268; 2.8217; 2.9044;
3.1746; 3.5742; 3.6312; 3.3957; 3.2943; 3.7544;
3.5478; 3.8203; 3.325; 3.8; 3.8352; 3.6312];
Y  = [242.0; 290.0; 340.0; 363.0; 430.0; 450.0;
500.0; 390.0; 450.0; 500.0; 475.0; 500.0; 500.0;
340.0; 600.0; 600.0; 700.0; 700.0; 610.0; 650.0;
```

```
575.0; 685.0; 620.0; 680.0; 700.0; 725.0; 720.0;
714.0; 850.0; 1000.0; 920.0; 955.0; 925.0; 975.0;
950.0; 40.0; 69.0; 78.0; 87.0; 120.0; 161.0; 110.0;
120.0; 150.0; 145.0; 160.0; 140.0; 160.0; 169.0;
161.0];

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];

Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);

% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
```

```
legend('Y aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

3.5.7. Penjelasan *Source Code*

1. Menyusun sistem normal equations (A, B)

```
%====== Multiple Regresi Linier ======
A = [50.0,  1352.4,  1473.8,  624.3352,  240.7445;
1352.4,  38342.52,  41809.28,  18164.31133,
6837.80626;
1473.8,  41809.28,  45593.56,  19822.34069,
7458.3868;
624.3352,  18164.31133,  19822.34069,  8785.203562,
3248.063845;
240.7445,  6837.80626,  7458.3868,  3248.063845,
1223.394672];

B = [23494.0; 716978.0; 782917.4; 353813.0909;
128681.0999];

AB = [A, B];
[n, ~] = size(AB);
```

Bagian ini memuat matriks normal equations hasil $X^T X$ (disimpan sebagai A) dan vektor $X^T y$ (disimpan sebagai B). Keduanya digabung menjadi matriks

teraugmentasi AB = [A B] berukuran 5×6. Nilai n = 5 adalah banyaknya koefisien yang ingin dicari: b_0, b_1, b_2, b_3, b_4.

2. Eliminasi Gauss dengan partial pivoting

```
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
```

Partial pivoting: memilih elemen terbesar (secara absolut) pada kolom ke-i di baris i..n sebagai pivot, lalu menukar baris bila perlu. Tujuannya menstabilkan perhitungan dan mengurangi pembagian dengan nilai kecil.

Eliminasi maju: untuk setiap baris di bawah pivot (indeks j), dihitung faktor Factor = a_{ji}/a_{ii} kemudian dikurangi baris j dengan factor × baris i. Hasilnya, elemen di bawah pivot menjadi nol dan terbentuk matriks segitiga atas.

3. Substitusi balik → mendapatkan koefisien *b*

```
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
```

```
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);
```

Setelah bentuk segitiga atas diperoleh, dilakukan back-substitution: untuk baris i dihitung jumlah kontribusi kolom kanan $s = \sum_{kk=i+1}^{n} a_{ik} b_{k`}$ lalu $b_i = \frac{a_{i,n+1} - s}{a_{ii}}$. Hasil akhirnya adalah vektor koefisien regresi $b = [b_0, \dots, b_4]^T$.

4. Menyusun data asli dan matriks desain *X*

```
X1 = [23.2; 24.0; 23.9; 26.3; 26.5; 26.8; 26.8;
27.6; 27.6; 28.5; 28.4; 28.7; 29.1; 29.5; 29.4;
29.4; 30.4; 30.4; 30.9; 31.0; 31.3; 31.4; 31.5;
31.8; 31.9; 31.8; 32.0; 32.7; 32.8; 33.5; 35.0;
35.0; 36.2; 37.4; 38.0; 12.9; 16.5; 17.5; 18.2;
18.6; 22.0; 19.1; 19.4; 20.4; 20.5; 20.5; 21.0;
21.1; 22.0; 22.0];
X2 = [25.4; 26.3; 26.5; 29.0; 29.0; 29.7; 29.7;
30.0; 30.0; 30.7; 31.0; 31.0; 31.5; 32.0; 32.0;
32.0; 33.0; 33.0; 33.5; 33.5; 34.0; 34.0; 34.5;
35.0; 35.0; 35.0; 35.0; 36.0; 36.0; 37.0; 38.5;
38.5; 39.5; 41.0; 41.0; 14.1; 18.2; 18.8; 19.8;
20.0; 23.4; 20.8; 21.0; 22.0; 22.0; 22.5; 22.5;
22.5; 24.0; 23.4];
X3 = [11.52; 12.48; 12.3778; 12.73; 12.444; 13.6024;
14.1795; 12.67; 14.0049; 14.2266; 14.2628; 14.3714;
13.7592; 13.9129; 14.9544; 15.438; 14.8604; 14.938;
15.633; 14.4738; 15.1285; 15.9936; 15.5227; 15.4686;
16.2405; 16.36; 16.3618; 16.517; 16.8896; 18.957;
```

```
18.0369; 18.084; 18.7542; 18.6354; 17.6235; 4.1472;
5.2983; 5.5756; 5.6166; 6.216; 6.9153; 6.1677;
6.1146; 5.8045; 6.6339; 7.0334; 6.55; 6.4; 7.5344;
6.9153];
X4 = [4.02; 4.3056; 4.6961; 4.4555; 5.134; 4.9274;
5.2785; 4.69; 4.8438; 4.9594; 5.1042; 4.8146; 4.368;
5.0728; 5.1708; 5.58; 5.2854; 5.1975; 5.1338;
5.7276; 5.5695; 5.3704; 5.2801; 6.1306; 5.589;
6.0532; 6.09; 5.8515; 6.1984; 6.603; 6.3063; 6.292;
6.7497; 6.7473; 6.3705; 2.268; 2.8217; 2.9044;
3.1746; 3.5742; 3.6312; 3.3957; 3.2943; 3.7544;
3.5478; 3.8203; 3.325; 3.8; 3.8352; 3.6312];
Y  = [242.0; 290.0; 340.0; 363.0; 430.0; 450.0;
500.0; 390.0; 450.0; 500.0; 475.0; 500.0; 500.0;
340.0; 600.0; 600.0; 700.0; 700.0; 610.0; 650.0;
575.0; 685.0; 620.0; 680.0; 700.0; 725.0; 720.0;
714.0; 850.0; 1000.0; 920.0; 955.0; 925.0; 975.0;
950.0; 40.0; 69.0; 78.0; 87.0; 120.0; 161.0; 110.0;
120.0; 150.0; 145.0; 160.0; 140.0; 160.0; 169.0;
161.0];
m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];
```

- X1 merepresentasikan Length 1 (vertical length) dalam sentimeter, yaitu panjang tubuh ikan dari mulut hingga pangkal ekor.
- X2 adalah Length 2 (diagonal length), yakni panjang diagonal dari mulut hingga ujung ekor.
- X3 menunjukkan Height, yaitu tinggi maksimum badan ikan. Nilai-nilainya berbeda tipis tetapi sangat mempengaruhi berat.

- X4 memuat Width, yaitu lebar ikan. Lebar ini juga merupakan variabel penting untuk memperkirakan volume tubuh ikan. Keempat vektor tersebut masing-masing berisi 50 baris karena dataset memiliki 50 sampel ikan.

Y adalah vektor berat ikan (Weight) dalam gram. Nilai-nilai di dalamnya berkisar dari sekitar 40 gram hingga 1 000 gram, sesuai dengan ukuran ikan pada observasi yang bersangkutan. Baris m = rows(Y); menghitung jumlah baris vektor Y. Variabel m ini mewakili banyaknya observasi dan digunakan untuk membangun kolom konstanta pada matriks desain. X = [ones(m,1), X1, X2, X3, X4]; membuat matriks berukuran m×5 dengan kolom pertama berisi angka 1 untuk koefisien intersep, diikuti empat kolom variabel bebas (X1–X4). Setiap baris matriks X berisi atribut lengkap satu ekor ikan. Matriks inilah yang kemudian digunakan dalam perhitungan regresi linier berganda untuk mendapatkan koefisien $b0$ – $b4$yang menggambarkan bagaimana panjang, lebar, dan tinggi ikan mempengaruhi beratnya.

5. Prediksi dan metrik galat

```
Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);
```

$\hat{Y} = Xb$ adalah prediksi, Residual e = Y - $\hat{Y}$, MSE = SSE dibagi jumlah observasi, RMSE = akar dari MSE → satuannya sama dengan Y, MAPE menyatakan rata-

rata persentase kesalahan absolut (dalam %). Penulisan %%%% pada sprintf/printf memang diperlukan untuk mencetak tanda % di output.

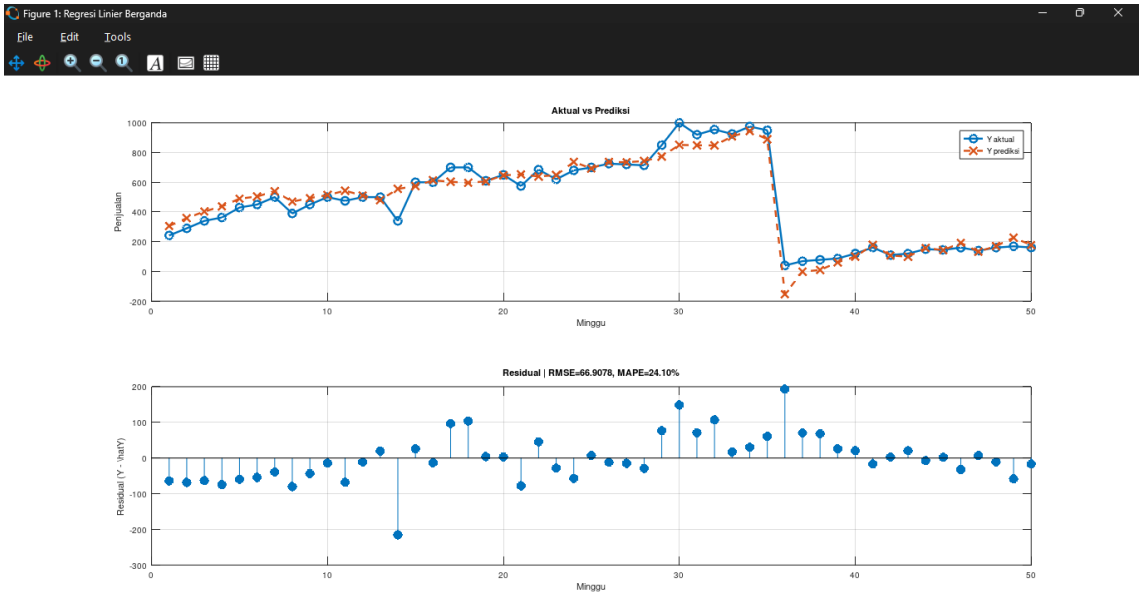
6. Grafik

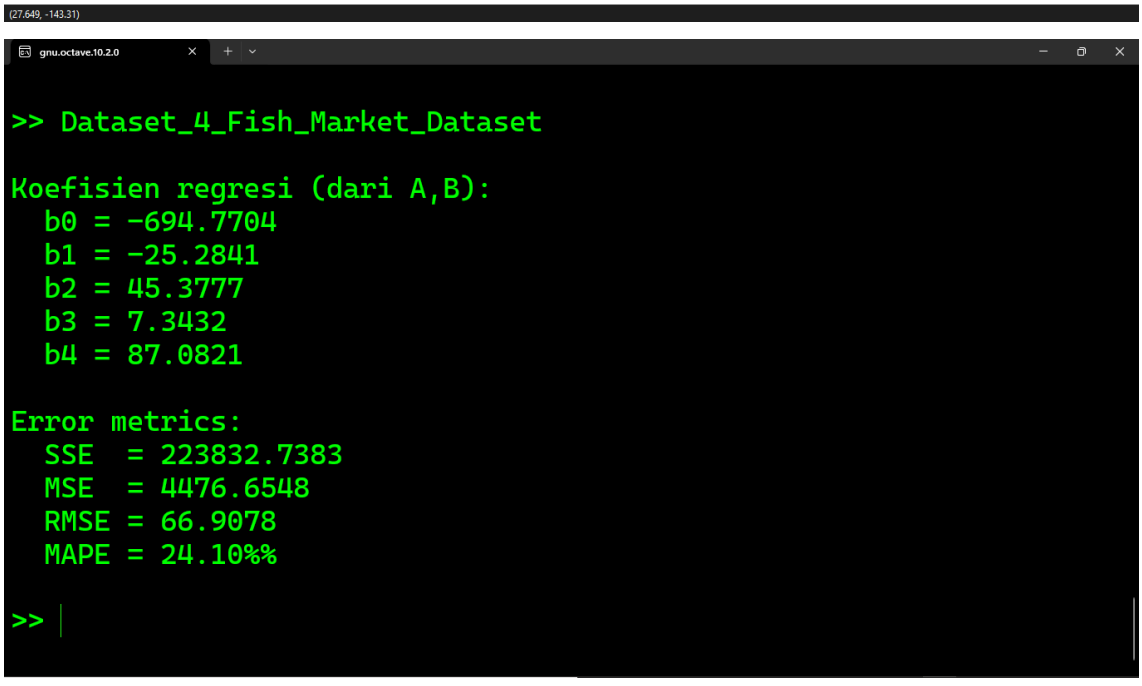
```
% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y                                    aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual  |  RMSE=%.4f,  MAPE=%.2f%%',
RMSE, MAPE));
```

- Subplot atas: membandingkan kurva aktual dan prediksi untuk melihat kedekatan pola.
- Subplot bawah: menampilkan residual per minggu. Titik yang jauh dari garis nol dapat diinterpretasi sebagai kandidat outlier atau observasi yang kurang terjelaskan oleh model.

  Di title(sprintf(...)) tutup format MAPE dengan %% agar tanda persen tercetak.

3.5.8. Tampilan Grafik dan Terminal





3.6. Implementasi Metode Multiple Regresi Linier Terhadap 3D Printer Dataset

3.6.1. Tahap Pengerjaan

Membentuk matriks desain X dengan ukuran $m \times (n+1)$, di mana m adalah jumlah observasi dan n jumlah variabel bebas. Kolom pertama diisi dengan angka 1 untuk merepresentasikan intersep, sedangkan kolom berikutnya diisi variabel bebas (X1–X4).

Menyusun vektor target Y dengan panjang m yang memuat variabel terikat pada setiap observasi.

Sigma (X1)	Sigma (X2)	Sigma (X3)	Sigma (X4)	Sigma (Y)	Sigma X1^2
5.3	2670	11075	3200.00	1004	0.765

Sigma (X2^2)	Sigma (X3^3)	Sigma (X4^2)
174100	2463875	248000

Sigma (X1X2)	Sigma (X1X3)	Sigma (X1X4)	Sigma (X2X3)	Sigma (X2X4)	Sigma (X3X4)
283.3	1173.95	334	595800	167400	708800

Sigma (X1Y)	Sigma (X2Y)	Sigma (X3Y)	Sigma (X4Y)
116	57590	219755	60820

3.6.2. Menyusun Matriks Augmented

$$\begin{bmatrix} N & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum (X1^2) & \sum (X2.X1) & \sum (X3.X1) & \sum (X4.X1) \\ \sum X2 & \sum (X1.X2) & \sum (X2^2) & \sum (X3.X2) & \sum (X4.X2) \\ \sum X3 & \sum (X1X3) & \sum (X2.X3) & \sum (X3^2) & \sum (X4.X3) \\ \sum X4 & \sum (X1.X4) & \sum (X2.X4) & \sum (X3.X4) & \sum (X4^2) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \sum (Y) \\ \sum (X1.Y) \\ \sum (X2.Y) \\ \sum (X3.Y) \\ \sum (X4.\ Y) \end{bmatrix}$$

Membentuk normal equation dengan menghitung $A = X^T X$ dan $B = X^T Y$. Matriks A berukuran $(n + 1) \times (n + 1)$ dan vektor B berukuran $(n + 1) \times 1$.

Matriks Augmented					
50	5.3	2670	11075	3200.00	1004
5.3	0.765	283.3	1173.95	334	116
2670	283.3	174100	595800	167400	57590

11075	1173.95	595800	2463875	708800	219755
3200.00	334	167400	708800	248000	60820

3.6.3. Tranformasi ke Bentuk Segitiga Atas

Eliminasi Gauss: Matriks augmented $[AB]$ direduksi menjadi bentuk segitiga atas menggunakan eliminasi baris elementer. Proses pivoting dilakukan untuk menghindari pembagian oleh angka mendekati nol. Penulis membagi seluruh baris pertama dengan 50 supaya elemen pertama menjadi 1.

$$R_1 \leftarrow \frac{1}{50} R_1$$

Matriks Bentuk Segitiga Atas Tahap 1 (R1 <- 1/50R1)					
1	0.106	53.4	221.5	64	20.08
5.3	0.765	283.3	1173.95	334	115.95
2670	283.3	174100	595800	167400	57590
11075	1173.95	595800	2463875	708800	219755
3200.00	334	167400	708800	248000	60820

Penulis menginginkan kolom pertama menjadi seperti:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Jadi penulis melakukan:

$$R_2 \leftarrow R_2 - 5.3R_1$$
$$R_3 \leftarrow R_3 - 2670R_1$$

$$R_4 \leftarrow R_4 - 11075R_1$$

$$R_5 \leftarrow R_5 - 3200R_1$$

Setelah eliminasi bentuknya menjadi:

Matriks Bentuk Segitiga Atas Tahap (R2 <-R2 - 5.3R1, R3 <-R3 - 2670R1, R4 <-R4 - 11075R1, R5 <-R5 - 3200R1)					
1	0.106	53.4	221.5	64	20.08
0	0.2032	0.28	0	-5.2	9.526
0	0.28	31522	4395	-3480	3976.4
0	0	4395	10762.5	0	-2631
0	-5.2	-3480	0	43200	-3436

Matriks Bentuk Segitiga Atas Tahap ($R_3 \leftarrow R_3 - 0.28R_2$, $R_4 \leftarrow R_4 - 0R_2$, $R_5 \leftarrow R_5 - (-5.2)R_2$)					
1	0.106	53.4	221.5	64	20.08
0	1	1.3780	0.0000	-25.5906	46.8799
0	0	31521.6142	4395.0000	-3472.8346	3963.2736
0	0	4395.0000	10762.5000	0.0000	-2631.0000
0	0	-3472.8346	0.0000	43066.9291	-3192.2244

Matriks Bentuk Segitiga Atas Tahap (Normalisasi)					
1	0.106	53.4	221.5	64	20.08
0	1	1.3780	0.0000	-25.5906	46.8799
0	0	1	0.1394	-0.1102	0.1257
0	0	0	10149.7133	484.2109	-3183.5919
0	0	0	484.2109	42684.3161	-2755.5782

Matriks Bentuk Segitiga Atas Tahap (Normalisasi R4)

1	0.106	53.4	221.5	64	20.08
0	1	1.3780	0.0000	-25.5906	46.8799
0	0	0	0.0000	0.0000	0.0000
0	0	0	1	0.0477	-0.3137
0	0	0	0	42661.2159	-2603.6990

Setelah matriks berada pada bentuk segitiga atas, koefisien regresi diperoleh b_0, b_1, b_2, b_3, b_4 dengan menyelesaikan sistem persamaan linear dari baris paling bawah ke atas.

Substitusi Mundur	
b0	79.3688
b1	45.0944
b2	0.1623
b3	-0.3108
b4	-0.0610

	Y Prediksi	e = Y - Y Prediksi
1	24.0743	-6.0743
2	22.5205	-6.5205
3	19.3434	-11.3434
4	14.6125	-4.6125
5	14.7517	-9.7517
6	22.1725	1.8275
7	15.7487	-3.7487
8	14.1949	-0.1949
9	22.3813	4.6187

10	15.9575	9.0425
11	23.0340	13.9660
12	11.7402	0.2598
13	15.0565	0.9435
14	5.4555	3.5445
15	8.8414	1.1586
16	30.8724	-3.8724
17	21.2019	1.7981
18	18.0248	7.9752
19	24.5878	8.4122
20	18.1640	10.8360
21	14.6825	1.3175
22	11.5054	0.4946
23	13.1983	-3.1983
24	14.9609	4.0391
25	8.6066	-0.6066
26	22.5209	-11.5209
27	19.3437	-7.3437
28	19.4133	-1.4133
29	24.3530	9.6470
30	13.0591	0.9409
31	22.2225	4.7775
32	14.1753	4.8247
33	19.1150	-1.1150
34	16.0074	-7.0074
35	17.7700	-4.7700

36	33.3076	-0.3076
37	26.8837	-2.8837
38	22.0833	3.9167
39	22.1529	-0.1529
40	15.7291	-11.7291
41	30.5679	4.4321
42	30.6375	3.3625
43	19.3436	8.6564
44	25.9762	2.0238
45	21.2453	-0.2453
46	33.5362	-5.5362
47	25.4890	-11.4890
48	30.4287	-0.4287
49	25.6282	3.3718
50	27.3212	-0.3212

3.6.4. Nilai Error

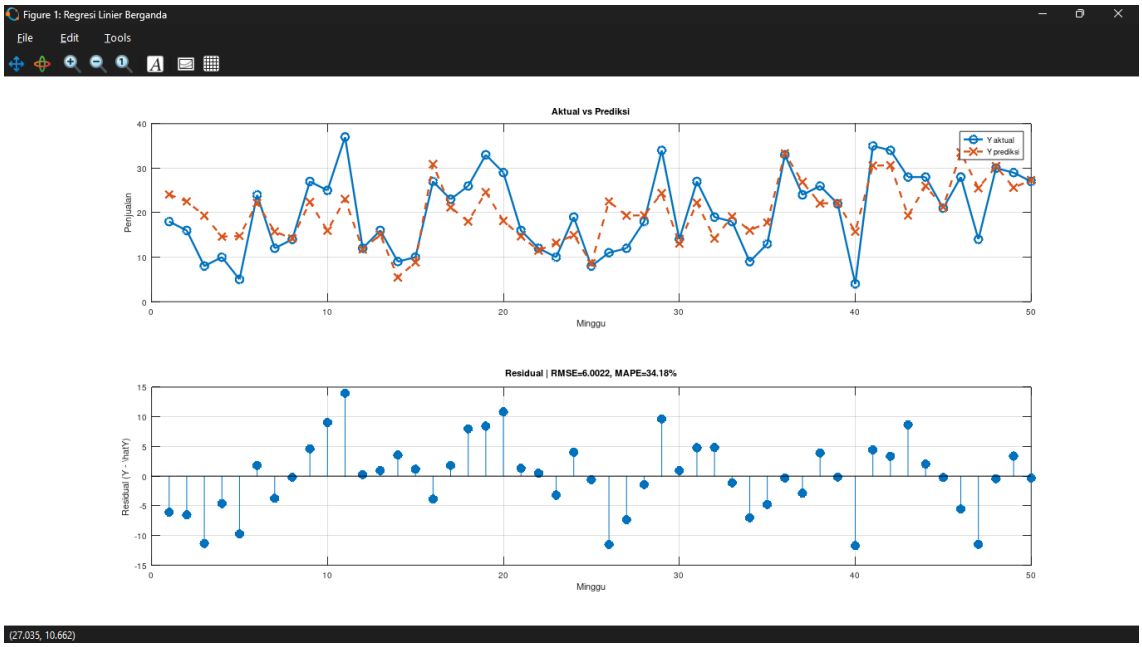
Vektor koefisien b digunakan untuk menghitung prediksi dengan rumus $\hat{Y} = Xb$,. Selisih antara nilai aktual dan prediksi Y - $\hat{Y}$ membentuk residual yang dianalisis untuk mengidentifikasi outlier.

RSS	1801.3072
MSE	36.0261
RMSE	6.0022
MAPE	34.1843

Deskripsi dataset menyatakan bahwa variabel masukan meliputi Layer Height (mm), Infill Density (%), Nozzle Temperature (°C) dan Print Speed (mm/s); variabel keluaran Tension Strength diukur dalam megapascal (MPa).

Hasil regresi memperlihatkan layer_height memiliki koefisien positif terbesar (~45,09 MPa per mm); artinya, peningkatan ketebalan lapisan 1 mm meningkatkan kekuatan tarik sekitar 45 MPa. Infill_density punya pengaruh positif kecil (~0,16 MPa per persen), sedangkan nozzle_temperature dan print_speed berkorelasi negatif (−0,31 MPa per °C dan −0,06 MPa per mm/s). Interpretasinya: lapisan yang lebih tebal dan kepadatan isi lebih tinggi membuat benda cetak lebih kuat, sedangkan suhu nozzle terlalu tinggi atau kecepatan cetak tinggi justru menurunkan kekuatan tarik, mungkin karena material tidak sempat mengendap sempurna.

3.6.5. Outlier



Dalam studi regresi linier berganda pada 3D Printer Dataset, outlier muncul sebagai hasil pengamatan percobaan yang menghasilkan kekuatan tarik (tension strength) jauh berbeda dari yang diprediksi berdasarkan parameter pencetakan. Variabel bebas yang dianalisis meliputi layer height, infill density, nozzle temperature, dan print speed. Nilai outlier biasanya disebabkan oleh kombinasi parameter yang menghasilkan produk

cetak dengan kualitas sangat baik atau sangat buruk, sehingga nilai kekuatan tarik berada di luar pola umum.

Pada Gambar, kurva biru merepresentasikan kekuatan tarik aktual dari setiap sampel, sedangkan tanda silang oranye menunjukkan prediksi model. Terlihat bahwa sebagian besar titik mengikuti tren yang berfluktuasi di kisaran 5–35 MPa, namun terdapat beberapa observasi yang menyimpang. Misalnya, sekitar percobaan ke-11 dan ke-19, nilai aktual melonjak mendekati 40 MPa sementara model memprediksi nilai yang lebih rendah, menghasilkan residual positif yang besar di grafik bawah. Sebaliknya, pada percobaan ke-40, model memprediksi kekuatan tarik lebih tinggi dibanding nilai aktual, sehingga residualnya bernilai negatif besar. Titik-titik residual yang berada jauh di atas atau di bawah garis nol inilah yang dikategorikan sebagai outlier.

Kehadiran outlier ini menunjukkan bahwa terdapat faktor lain (misalnya jenis material atau variasi peralatan) yang mempengaruhi kekuatan tarik tetapi tidak dimasukkan ke dalam model, atau adanya ketidakstabilan proses pencetakan yang menyebabkan hasil tidak konsisten. Mengenali dan memahami outlier membantu mahasiswa untuk mengevaluasi kelayakan model linier dan mempertimbangkan transformasi variabel atau tambahan faktor agar prediksi lebih akurat.

3.6.6. *Source Code*

```
%====== Multiple Regresi Linier ======
A = [50.0,  5.3,  2670.0,  11075.0,  3200.0;
5.3,  0.765,  283.3,  1173.95,  334.0;
2670.0,  283.3,  174100.0,  595800.0,  167400.0;
11075.0,  1173.95,  595800.0,  2463875.0,  708800.0;
3200.0,  334.0,  167400.0,  708800.0,  248000.0];

B = [1004.0; 115.95; 57590.0; 219755.0; 60820.0];

AB = [A, B];
```

```
[n, ~] = size(AB);
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);

X1 = [0.02; 0.02; 0.02; 0.02; 0.02; 0.02; 0.02;
0.02; 0.02; 0.02; 0.06; 0.06; 0.06; 0.06; 0.06;
0.06; 0.06; 0.06; 0.06; 0.06; 0.1; 0.1; 0.1; 0.1;
0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.15; 0.15; 0.15;
```

```
0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.2; 0.2;
0.2; 0.2; 0.2; 0.2; 0.2; 0.2; 0.2; 0.2];
X2 = [90.0; 90.0; 80.0; 70.0; 90.0; 40.0; 10.0;
10.0; 70.0; 40.0; 80.0; 20.0; 50.0; 10.0; 50.0;
90.0; 40.0; 30.0; 80.0; 50.0; 40.0; 30.0; 50.0;
80.0; 60.0; 50.0; 40.0; 50.0; 90.0; 30.0; 50.0;
10.0; 50.0; 50.0; 80.0; 80.0; 50.0; 30.0; 40.0;
10.0; 80.0; 90.0; 30.0; 90.0; 80.0; 60.0; 20.0;
60.0; 40.0; 60.0];
X3 = [220.0; 225.0; 230.0; 240.0; 250.0; 200.0;
205.0; 210.0; 215.0; 220.0; 220.0; 225.0; 230.0;
240.0; 250.0; 200.0; 205.0; 210.0; 215.0; 220.0;
220.0; 225.0; 230.0; 240.0; 250.0; 200.0; 205.0;
210.0; 215.0; 220.0; 220.0; 225.0; 230.0; 240.0;
250.0; 200.0; 205.0; 210.0; 215.0; 220.0; 220.0;
225.0; 230.0; 240.0; 250.0; 200.0; 205.0; 210.0;
215.0; 220.0];
X4 = [40.0; 40.0; 40.0; 40.0; 40.0; 40.0; 40.0;
40.0; 40.0; 40.0; 60.0; 60.0; 60.0; 60.0; 60.0;
60.0; 60.0; 60.0; 60.0; 60.0; 120.0; 120.0; 120.0;
120.0; 120.0; 120.0; 120.0; 120.0; 120.0; 120.0;
60.0; 60.0; 60.0; 60.0; 60.0; 60.0; 60.0; 60.0;
60.0; 60.0; 40.0; 40.0; 40.0; 40.0; 40.0; 40.0;
40.0; 40.0; 40.0; 40.0];
Y  = [18.0; 16.0; 8.0; 10.0; 5.0; 24.0; 12.0; 14.0;
27.0; 25.0; 37.0; 12.0; 16.0; 9.0; 10.0; 27.0; 23.0;
26.0; 33.0; 29.0; 16.0; 12.0; 10.0; 19.0; 8.0; 11.0;
12.0; 18.0; 34.0; 14.0; 27.0; 19.0; 18.0; 9.0; 13.0;
```

```
33.0; 24.0; 26.0; 22.0; 4.0; 35.0; 34.0; 28.0; 28.0;
21.0; 28.0; 14.0; 30.0; 29.0; 27.0];

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];

Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);

% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
title('Aktual vs Prediksi');
legend('Y aktual','Y
prediksi','Location','northeast');
```

```
subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

3.6.7. Penjelasan *Source Code*

1. Menyusun sistem normal equations (A, B)

```
%====== Multiple Regresi Linier ======
A = [50.0,  5.3,  2670.0,  11075.0,  3200.0;
5.3,  0.765,  283.3,  1173.95,  334.0;
2670.0,  283.3,  174100.0,  595800.0,  167400.0;
11075.0,  1173.95,  595800.0,  2463875.0,  708800.0;
3200.0,  334.0,  167400.0,  708800.0,  248000.0];

B = [1004.0; 115.95; 57590.0; 219755.0; 60820.0];

AB = [A, B];
[n, ~] = size(AB);
```

Bagian ini memuat matriks normal equations hasil $X^T X$ (disimpan sebagai A) dan vektor $X^T y$ (disimpan sebagai B). Keduanya digabung menjadi matriks teraugmentasi AB = [A B] berukuran 5×6. Nilai n = 5 adalah banyaknya koefisien yang ingin dicari: b_0, b_1, b_2, b_3, b_4.

2. Eliminasi Gauss dengan partial pivoting

```
for i = 1:n
  [~, pivotIndex] = max(abs(AB(i:n, i)));
  pivotIndex = pivotIndex + (i - 1);
  if pivotIndex ~= i
    AB([i, pivotIndex], :) = AB([pivotIndex, i], :);
```

```
  end
  for j = i + 1:n
    factor = AB(j, i) / AB(i, i);
    AB(j, i:end) = AB(j, i:end) - factor * AB(i,
i:end);
  end
end
```

Partial pivoting: memilih elemen terbesar (secara absolut) pada kolom ke-i di baris i..n sebagai pivot, lalu menukar baris bila perlu. Tujuannya menstabilkan perhitungan dan mengurangi pembagian dengan nilai kecil.

Eliminasi maju: untuk setiap baris di bawah pivot (indeks j), dihitung faktor Factor = a_{ji}/a_{ii} kemudian dikurangi baris j dengan factor × baris i. Hasilnya, elemen di bawah pivot menjadi nol dan terbentuk matriks segitiga atas.

3. Substitusi balik → mendapatkan koefisien *b*

```
b = zeros(n,1);
for i = n:-1:1
  s = AB(i, i+1:n)* b(i+1:n);
  b(i) = (AB(i, end)- s)/AB(i, i);
end
b0 = b(1); b1 = b(2); b2 = b(3); b3 = b(4); b4 =
b(5);

printf('Koefisien regresi (dari A,B):\n');
printf('  b0 = %.4f\n  b1 = %.4f\n  b2 = %.4f\n  b3
= %.4f\n  b4 = %.4f\n\n', b0,b1,b2,b3,b4);
```

Setelah bentuk segitiga atas diperoleh, dilakukan back-substitution: untuk baris i dihitung jumlah kontribusi kolom kanan $s = \sum_{kk=i+1}^{n} a_{ik} b_{k`}$ lalu $b_i = \frac{a_{i,n+1} - s}{a_{ii}}$. Hasil akhirnya adalah vektor koefisien regresi $b = [b_0, \ldots, b_4]^T$.

4. Menyusun data asli dan matriks desain *X*

```
X1 = [0.02; 0.02; 0.02; 0.02; 0.02; 0.02; 0.02;
0.02; 0.02; 0.02; 0.06; 0.06; 0.06; 0.06; 0.06;
0.06; 0.06; 0.06; 0.06; 0.06; 0.1; 0.1; 0.1; 0.1;
0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.15; 0.15; 0.15;
0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.2; 0.2;
0.2; 0.2; 0.2; 0.2; 0.2; 0.2; 0.2; 0.2];
X2 = [90.0; 90.0; 80.0; 70.0; 90.0; 40.0; 10.0;
10.0; 70.0; 40.0; 80.0; 20.0; 50.0; 10.0; 50.0;
90.0; 40.0; 30.0; 80.0; 50.0; 40.0; 30.0; 50.0;
80.0; 60.0; 50.0; 40.0; 50.0; 90.0; 30.0; 50.0;
10.0; 50.0; 50.0; 80.0; 80.0; 50.0; 30.0; 40.0;
10.0; 80.0; 90.0; 30.0; 90.0; 80.0; 60.0; 20.0;
60.0; 40.0; 60.0];
X3 = [220.0; 225.0; 230.0; 240.0; 250.0; 200.0;
205.0; 210.0; 215.0; 220.0; 220.0; 225.0; 230.0;
240.0; 250.0; 200.0; 205.0; 210.0; 215.0; 220.0;
220.0; 225.0; 230.0; 240.0; 250.0; 200.0; 205.0;
210.0; 215.0; 220.0; 220.0; 225.0; 230.0; 240.0;
250.0; 200.0; 205.0; 210.0; 215.0; 220.0; 220.0;
225.0; 230.0; 240.0; 250.0; 200.0; 205.0; 210.0;
215.0; 220.0];
X4 = [40.0; 40.0; 40.0; 40.0; 40.0; 40.0; 40.0;
40.0; 40.0; 40.0; 60.0; 60.0; 60.0; 60.0; 60.0;
60.0; 60.0; 60.0; 60.0; 60.0; 120.0; 120.0; 120.0;
120.0; 120.0; 120.0; 120.0; 120.0; 120.0; 120.0;
```

```
60.0; 60.0; 60.0; 60.0; 60.0; 60.0; 60.0; 60.0;
60.0; 60.0; 40.0; 40.0; 40.0; 40.0; 40.0; 40.0;
40.0; 40.0; 40.0; 40.0];
Y  = [18.0; 16.0; 8.0; 10.0; 5.0; 24.0; 12.0; 14.0;
27.0; 25.0; 37.0; 12.0; 16.0; 9.0; 10.0; 27.0; 23.0;
26.0; 33.0; 29.0; 16.0; 12.0; 10.0; 19.0; 8.0; 11.0;
12.0; 18.0; 34.0; 14.0; 27.0; 19.0; 18.0; 9.0; 13.0;
33.0; 24.0; 26.0; 22.0; 4.0; 35.0; 34.0; 28.0; 28.0;
21.0; 28.0; 14.0; 30.0; 29.0; 27.0];

m  = rows(Y);
X  = [ones(m,1), X1, X2, X3, X4];
```

- X1 menyimpan nilai layer height (tinggi lapisan) dalam milimeter. Variasi tinggi lapisan (0,02; 0,06; 0,10; 0,15; 0,20) berpengaruh terhadap waktu pencetakan dan kekuatan benda cetak.
- X2 adalah infill density dalam persen, yaitu kepadatan isi bagian dalam objek (10–90 %). Kepadatan yang lebih tinggi biasanya menghasilkan produk lebih kokoh.
- X3 memuat nozzle temperature (suhu nozzle) dalam derajat Celsius. Suhu bervariasi dari 200 °C hingga 250 °C dan mempengaruhi peleburan material.
- X4 berisi print speed (kecepatan cetak) dalam mm/s, dengan nilai 40, 60, dan 120. Kecepatan cetak yang berbeda memberikan waktu pendinginan yang berbeda sehingga mempengaruhi kekuatan tarik.

Vektor Y berisi data tension strength atau kekuatan tarik hasil cetakan (dalam MPa). Nilai-nilainya berkisar dari 4 hingga hampir 40, mencerminkan variasi kualitas hasil cetak pada setiap konfigurasi percobaan. m = rows(Y); menghitung jumlah baris pada vektor Y, yang merupakan banyaknya percobaan (50). Nilai m digunakan untuk membangun kolom konstanta. X = [ones(m,1), X1, X2, X3, X4]; membentuk matriks desain berukuran m×5. Kolom pertama berisi angka 1

sebagai konstanta untuk menghitung intercept $b0$. Empat kolom berikutnya adalah variabel layer height, infill density, nozzle temperature, dan print speed. Dengan format ini, setiap baris matriks X memuat parameter lengkap dari satu percobaan. Matriks X kemudian digunakan bersama vektor Y dalam persamaan normal untuk memperoleh koefisien regresi $b0 – b4$, yang menjelaskan kontribusi masing-masing parameter terhadap kekuatan tarik hasil cetakan.

5. Prediksi dan metrik galat

```
Yprediksi = X * b;
e    = Y - Yprediksi;
SSE  = sum(e.^2);
MSE  = SSE / m;
RMSE = sqrt(MSE);
MAPE = mean(abs(e) ./ abs(Y)) * 100;

printf('Error metrics:\n');
printf('  SSE  = %.4f\n', SSE);
printf('  MSE  = %.4f\n', MSE);
printf('  RMSE = %.4f\n', RMSE);
printf('  MAPE = %.2f%%%%\n\n', MAPE);
```

$\hat{Y} = Xb$ adalah prediksi, Residual e = Y - $\hat{Y}$, MSE = SSE dibagi jumlah observasi, RMSE = akar dari MSE → satuannya sama dengan Y, MAPE menyatakan rata-rata persentase kesalahan absolut (dalam %). Penulisan %%%% pada sprintf/printf memang diperlukan untuk mencetak tanda % di output.

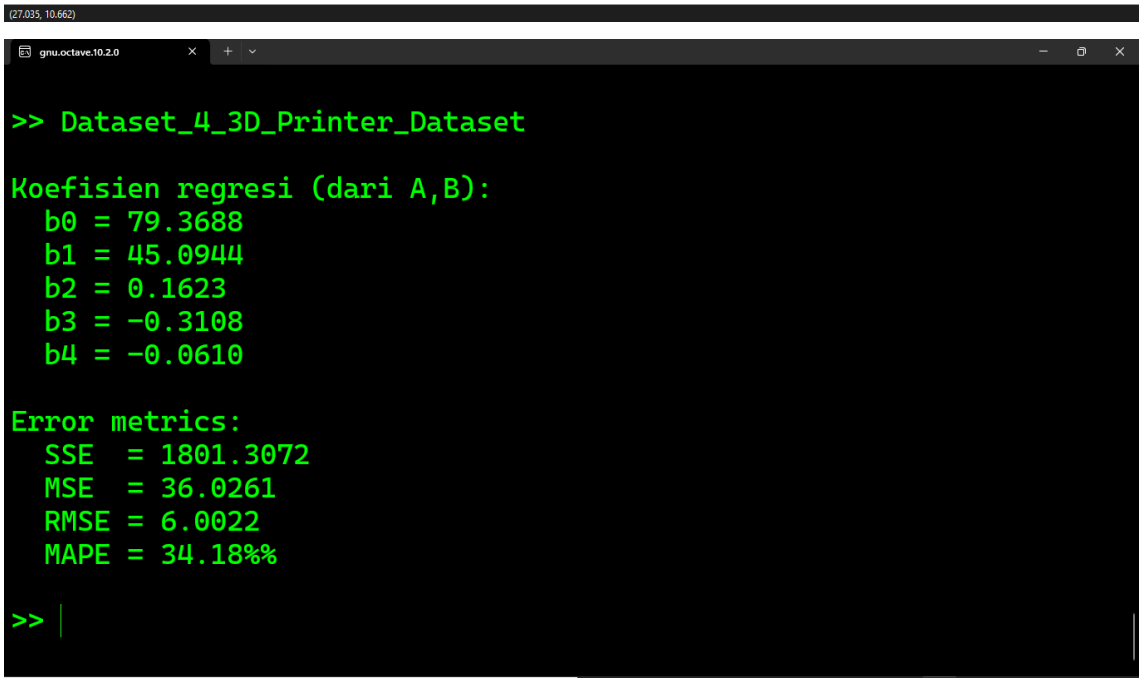
6. Grafik

```
% ====== Setting Grafik ======
figure('Name','Regresi Linier Berganda');
subplot(2,1,1);
plot(1:m, Y, 'o-','LineWidth',1.5); hold on;
plot(1:m, Yprediksi, 'x--','LineWidth',1.5);
grid on; xlabel('Minggu'); ylabel('Penjualan');
```

```
title('Aktual vs Prediksi');
legend('Y                                    aktual','Y
prediksi','Location','northeast');

subplot(2,1,2);
stem(1:m, e, 'filled'); grid on;
xlabel('Minggu'); ylabel('Residual (Y - \hat{Y})');
% atau 'Residual (Y - Ypred)'
title(sprintf('Residual | RMSE=%.4f, MAPE=%.2f%%',
RMSE, MAPE));
```

- Subplot atas: membandingkan kurva aktual dan prediksi untuk melihat kedekatan pola.
- Subplot bawah: menampilkan residual per minggu. Titik yang jauh dari garis nol dapat diinterpretasi sebagai kandidat outlier atau observasi yang kurang terjelaskan oleh model.
  Di title(sprintf(...)) tutup format MAPE dengan %% agar tanda persen tercetak.

3.6.8. Tampilan Grafik dan Terminal



3.7. Evaluasi Performa Model

Dataset	RSS	MSE	RMSE	MAPE	Interpretasi
Housing Prices	71241851299203	1424837025984.06	1193665.3744	9.29%	Harga rumah bernilai jutaan, sehingga RMSE sekitar 1,19 juta masih cukup

					baik, MAPE di bawah 10 % menandakan prediksi relatif akurat.
Car Price	487208826.4937	9744176.5299	3121.5664	19.37%	RMSE sekitar 3 121 dolar dan MAPE hampir 20 % menunjukkan variasi harga mobil lebih sulit ditangkap oleh model linier sederhana.
Medical Cost Personal	88651670924548	1773033418490.96	1331553.0100	11.10%	Premi asuransi berada pada kisaran 7–13 juta; RMSE 1,33 juta cukup tinggi tetapi MAPE ~11 % masih dapat diterima.
Fish Market	223832.7383	4476.6548	66.9078	24.10%	Prediksi berat ikan memiliki RMSE ±67 g; MAPE 24 % menunjukkan adanya pengaruh nonlinier atau variabel lain yang belum dipertimbangkan.
3D Printer	1801.3072	36.0261	6.0022	34.18%	RMSE kecil (6 MPa) karena nilai target

					berkisar 4–40 MPa; MAPE tinggi (34 %) mengindikasikan fluktuasi besar dan potensi model nonlinier lebih sesuai.

BAB IV
PENUTUP

4.1. Kesimpulan

Berdasarkan hasil implementasi dan analisis yang dilakukan, dapat disimpulkan bahwa:

- Matriks normal equation yang disusun dari variabel bebas dan variabel terikat masing-masing dataset berhasil diolah menjadi matriks segitiga atas dan selanjutnya diselesaikan menggunakan back-substitution. Koefisien regresi yang diperoleh menunjukkan kontribusi relatif setiap variabel bebas terhadap variabel terikat.
- Nilai SSE, MSE, RMSE, dan MAPE yang dihitung memberikan gambaran kuantitatif tentang akurasi model. Secara umum, model mampu mengikuti tren data dengan cukup baik, namun akurasi sangat bergantung pada kualitas dan karakteristik dataset. Dataset dengan variasi variabel bebas yang jelas dan hubungan linier yang kuat, seperti Housing Prices dan Fish Market, menghasilkan nilai error lebih kecil dibanding dataset yang memiliki pola lebih kompleks, seperti 3D Printer.
- Grafika residual memudahkan identifikasi observasi yang menyimpang jauh dari prediksi. Kehadiran outlier terbukti meningkatkan nilai error dan menurunkan MAPE. Pada beberapa dataset, outlier disebabkan oleh kombinasi variabel yang ekstrem atau faktor yang tidak dimodelkan, sehingga penting bagi analis untuk mempertimbangkan transformasi data atau penggunaan model yang lebih fleksibel.
- Model linier berganda hanya mampu menangkap hubungan linier antar variabel. Pada dataset dengan hubungan non-linier atau interaksi antar variabel yang kompleks, model linier menghasilkan error yang cukup besar. Hal ini menandakan perlunya eksplorasi metode lain seperti regresi polinomial atau algoritma machine learning.
- Housing Prices: Model regresi linier berganda mampu memprediksi harga rumah dengan akurasi cukup baik. Luas bangunan (area) adalah prediktor paling dominan, sedangkan jumlah kamar tidur dan kamar mandi memberikan kontribusi

kecil. Outlier umumnya berasal dari rumah dengan harga jauh di atas rata-rata, misalnya rumah mewah dengan fasilitas lengkap.

- Car Price: Dimensi fisik mobil dan tenaga mesin mempengaruhi harga, tetapi masih banyak variabel lain (merk, tahun produksi, fitur interior) yang tidak disertakan. Hal ini menyebabkan MAPE cukup tinggi. Pengembangan model sebaiknya memasukkan variabel tambahan untuk meningkatkan akurasi.
- Medical Cost: Usia, BMI, jumlah anak, dan status perokok terbukti mempengaruhi besarnya premi. Koefisien positif untuk smoker menegaskan bahwa perokok membayar premi lebih tinggi. Meskipun model cukup baik, variabel lain seperti jenis kelamin atau wilayah tinggal mungkin menambah keakuratan.
- Fish Market: Panjang, tinggi, dan lebar ikan memiliki korelasi dengan berat, tetapi hubungan tampaknya tidak sepenuhnya linier. Beberapa ikan dengan ukuran serupa tetapi berat berbeda menjadi outlier, menunjukkan adanya perbedaan spesies atau kepadatan daging yang tidak diakomodasi model.
- 3D Printer: Parameter tinggi lapisan, kepadatan isi, temperatur nozzle, dan kecepatan cetak berpengaruh terhadap kekuatan tarik cetakan. Namun hubungan yang kompleks dan interaksi antar variabel membuat model linier kurang akurat. Dibutuhkan pendekatan nonlinier atau penggunaan parameter lain seperti jenis material untuk prediksi yang lebih baik.

4.2. Saran

Beberapa saran untuk pengembangan analisis ke depan:

- Lakukan pembersihan data yang lebih teliti, termasuk penanganan missing value dan penghapusan outlier yang tidak representatif, agar model dapat belajar dari pola yang lebih konsisten.
- Pertimbangkan penambahan variabel bebas baru atau transformasi variabel (misalnya logaritma, pangkat) untuk menangkap pola non-linier dan interaksi antar variabel.
- Coba bandingkan performa regresi linier berganda dengan model lain seperti regresi ridge/lasso, regresi polinomial, atau algoritma machine learning (random

forest, gradient boosting), sehingga diperoleh model yang paling tepat untuk masing-masing dataset.

- Terapkan teknik validasi silang (cross-validation) untuk menguji keandalan model pada data yang belum pernah dilihat sebelumnya. Hal ini akan memberikan estimasi performa yang lebih realistis dan menghindari overfitting.
- Kembangkan visualisasi interaktif untuk membantu memonitor perubahan model saat variabel dan parameter diubah, sehingga proses analisis menjadi lebih intuitif.

Dengan demikian, laporan ini diharapkan tidak hanya menjadi dokumentasi pelaksanaan praktikum, tetapi juga memberikan gambaran menyeluruh tentang penerapan regresi linier berganda pada data nyata dan membuka wawasan untuk penelitian lanjutan.

DAFTAR PUSTAKA

- Farhanna Mar'i. (2025). *SK-5 Multiple Linear Regression* (Bahan Ajar Sains Komputasi). Surabaya: Universitas Negeri Surabaya.
- **Afumetto, A.** (2023). *3D Printer Dataset for Mechanical Engineers*. Kaggle. Tersedia secara daring di: https://www.kaggle.com/datasets/afumetto/3dprinter (diakses 8 Oktober 2025).
- **Choeda, N.** (2020). *Car Price Dataset*. Kaggle. Tersedia secara daring di: https://www.kaggle.com/datasets/ngawangchoeda/car-price-dataset (diakses 8 Oktober 2025).
- **Freedman, D. A.** (2009). *Statistical Models: Theory and Practice* (edisi ke-2). Cambridge: Cambridge University Press.
- **Gujarati, D. N. & Porter, D. C.** (2010). *Basic Econometrics* (edisi ke-5). New York: McGraw-Hill Education.
- **Mirichoi, M.** (2017). *Insurance – Medical Cost Personal Datasets*. Kaggle. Tersedia secara daring di: https://www.kaggle.com/datasets/mirichoi0218/insurance (diakses 8 Oktober 2025).
- **Statlect** (2022). *Multiple Linear Regression*. Tersedia secara daring di: https://www.statlect.com/fundamentals-of-statistics/multiple-linear-regression (diakses 8 Oktober 2025).
- **Yasser Hajji** (2021). *Housing Prices Dataset*. Kaggle. Tersedia secara daring di: https://www.kaggle.com/datasets/yasserh/housing-prices-dataset (diakses 8 Oktober 2025).
- **Kaggle Datasets Team** (2019). *Fish Market Dataset*. Kaggle. Tersedia secara daring di: https://www.kaggle.com/datasets/abbasi1214/fish-market-dataset (diakses 8 Oktober 2025).
- **MathWorks** (2020). *MATLAB and Octave Documentation: Linear Regression*. MathWorks Documentation. Tersedia secara daring di: https://www.mathworks.com/help/stats/linear-regression.html (diakses 8 Oktober 2025).

- **GitHub README – Predictive Analysis of Insurance Charges** (2023). "Medical costs billed by health insurance (USD)". Diakses dari https://raw.githubusercontent.com/thaochu05/Multiple-regression-Predicting-Insurance-Charges/master/README.md pada 8 Oktober 2025.
- **Fish Market Dataset** – HuggingFace. Kartu dataset menyatakan bahwa Weight diukur dalam gram dan Length1, Length2, Length3, Height, Width diukur dalam sentimeter.
- **APMonitor.com** (2023). *3D Printing Parameter Prediction*. Menjelaskan bahwa parameter layer height dan wall thickness diukur dalam mm, infill density dalam %, nozzle temperature dan bed temperature dalam °C, print speed dalam mm/s, serta keluaran tension strength dalam MPa.
- **Predictive Analysis of Insurance Charges** (GitHub). Menyajikan ringkasan statistik yang menyebutkan bahwa distribusi variabel charges berkisar antara $1 121,87 hingga $63 770,43.